АВТОМАТИКА
и
ТЕЛЕМЕХАНИКА

Cat

Volume 19, No. 1

January 1958

SOVIET INSTRUMENTATION AND
CONTROL TRANSLATION SERIES

Automation and Remote Control

(The Soviet Journal *Avtomatika i Telemekhanika* in English Translation)

■ **This translation of a Soviet journal on automatic control is published as a service to American science and industry. It is sponsored by the Instrument Society of America under a grant in aid from the National Science Foundation, continuing a program initiated by the Massachusetts Institute of Technology.**

SOVIET INSTRUMENTATION AND CONTROL TRANSLATION SERIES

■ This translation of the Soviet Journal *Avtomatika i Telemekhanika* is published and distributed at nominal subscription rates under a grant in aid to the Instrument Society of America from the National Science Foundation. This translated journal, and others in the Series (see back cover), will enable American scientists and engineers to be informed of work in the fields of instrumentation, measurement techniques and automatic control reported in the Soviet Union.

The original Russian articles are translated by competent technical personnel. The translations are on a cover-to-cover basis, permitting readers to appraise for themselves the scope, status and importance of the Soviet work.

Publication of *Avtomatika i Telemekhanika* in English translation started under the present auspices in April 1958 with Russian Vol. 18, No. 1 of January 1957. Translation of Vol. 18 has now been completed. The twelve issues of Vol. 19 will be published in English translation by mid-1959.

All views expressed in the translated material are intended to be those of the original authors, and not those of the translators, nor the Instrument Society of America.

Readers are invited to submit communications on the quality of the translations and the content of the articles to ISA headquarters. Pertinent correspondence will be published in the "Letters" section of the ISA Journal. Space will also be made available in the ISA Journal for such replies as may be received from Russian authors to comments or questions by American readers.

Subscription Prices:

Per year (12 issues), starting with Vol. 19, No. 1

General: United States and Canada	$30.00
Elsewhere	33.00
Libraries of non-profit academic institutions:	
United States and Canada	$15.00
Elsewhere	18.00
Single issues to everyone, each	$ 6.00

See back cover for combined subscription to entire Series.

Subscriptions and requests for information on back issues should be addressed to the:

Instrument Society of America
313 Sixth Avenue, Pittsburgh 22, Penna.

Translated and printed by Consultants Bureau, Inc.

Volume XIX No. 1 – January 1958
English Translation Printed January 1959

Automation and Remote Control

The Soviet Journal Avtomatika i Telemekhanika in English Translation

Avtomatika i Telemekhanika is a Publication of the Academy of Sciences of the USSR

See following page for Table of Contents.

AUTOMATION AND REMOTE CONTROL

VOLUME 19, NUMBER 1 JANUARY, 1958

CONTENTS

RELAXING THE SUFFICIENCY CONDITIONS FOR ABSOLUTE STABILITY

Vasile-Mikhai Popov

(Power Institute of the Academy Science of the Rumanian People's Republic)

The sufficiency conditions for absolute stability are investigated in the case of automatic control systems which contain a servomotor with a nonlinear speed characteristic. It is shown that it is possible to relax these conditions in certain cases. In particular, necessary and sufficient conditions are obtained for three cases (except for intractable special cases).

We consider an automatic control system described by the equations

$$\dot{\eta}_k = \sum_{\alpha=1}^{n} b_{k\alpha}\eta_\alpha + n_k\xi \quad (k = 1, \ldots, n),$$
$$\dot{\xi} = f(\sigma), \quad \sigma = \sum_{\alpha=1}^{n} p_\alpha \eta_\alpha - \xi.$$

where $b_{k\alpha}$ are constant parameters, η_α are the coordinates of the controlled object, η_k are constant parameters, ξ is the coordinate of the controlling organ, p_α are controller parameters, $f(\sigma)$ is a continuous function, having only a finite number of first-type discontinuities, and with the properties $f(0) = 0$ and $\sigma f(\sigma) > 0$ [class (A)].

In [1] there is reproduced the proof of a theorem due to A. I. Lur'e [2], determining the existence of criteria which, if satisfied, guarantee the absolute stability of a self-regulating system for any class (A) function $f(\sigma)$ and, consequently, for any linear function $f(\sigma) = h\sigma (0 < h < \infty)$.

Also in [1] (page 172) there was formulated the problem of inverting this theorem, i.e., the problem of ascertaining the validity of its converse, "that is, if a system is stable for an arbitrary function, $f_k(\sigma) = h\sigma (0 < h < \infty)$, then would it be correct to assert that the system would be stable for any class (A) function ?".

This problem is solved in the case of three equations. The solution is based on a lemma, to be proved below, which allows the extension, in parameter space, of the domain of absolute stability, obtained beforehand by applying Lur'e's algorithm [2].

1. Auxiliary Lemma

We consider an indirectly-controlled system described by the equations

$$\dot{z}_\rho = \lambda_\rho z_\rho + f(\sigma) \quad (\rho = 1, 2, \ldots, n), \tag{I}$$
$$\dot{\sigma} = \sum_{\rho=1}^{n} \beta_\rho z_\rho - rf(\sigma),$$

with the usual assumptions [1, 2].

The following lemma is asserted: if the trivial solution to the system

$$\dot{z}_\rho = \lambda_\rho z_\rho + f(\sigma) \quad (\rho = 1, 2, \ldots, n),$$
$$\dot{\sigma} = \sum_{\rho=1}^{n} \beta_\rho \left(1 + \frac{p}{\lambda_\rho}\right) z_\rho - r f(\sigma), \tag{II}$$

(where p is an arbitrary positive constant) possesses asymptotic absolute stability, and if this fact can be established by means of a Liapunov function of the form

$$v = -\Phi(z_1, z_2, \ldots, z_n) - \int_0^\sigma f(\sigma)\, d\sigma, \tag{1.1}$$

(where Φ is a positive definite quadratic form)and, moreover, if the following inequality holds:

$$r + \sum_{\rho=1}^{n} \frac{\beta_\rho}{\lambda_\rho} > 0, \tag{1.2}$$

then the trivial solution of System (I) also possesses asymptotic absolute stability.

We proved this lemma by constructing the Liapunov function for System (I).

From the assumptions stated above, the total derivative of the Function (1.1), taking Equations (II) into account,

$$\left(\frac{dv}{dt}\right)_{II} = -\frac{d\Phi(z_1, z_2, \ldots, z_n)}{dt} - f(\sigma)\left(\sum_{\rho=1}^{n} \beta_\rho z_\rho - r f(\sigma)\right) - p f(\sigma) \sum_{\rho=1}^{n} \frac{\beta_\rho}{\lambda_\rho} z_\rho \tag{1.3}$$

is a positive definite function.

We now prove that the function

$$v' = v - \frac{p}{2\left(r + \sum_{\rho=1}^{n} \frac{\beta_\rho}{\lambda_\rho}\right)} \left(\sum_{\rho=1}^{n} \frac{\beta_\rho}{\lambda_\rho} z_\rho - \sigma\right)^2, \tag{1.4}$$

which, by virtue of the assumptions already made, is negative definite, is indeed the Liapunov function for System (I). Indeed, the total derivative of this function, taking Equations (I) into account, has the form:

$$\left(\frac{dv'}{dt}\right)_I = -\frac{d\Phi(z_1, z_2, \ldots, z_n)}{dt} - f(\sigma)\left[\Sigma \beta_\rho z_\rho - r f(\sigma)\right] -$$
$$- p\left(\sum \frac{\beta_\rho}{\lambda_\rho} z_\rho - \sigma\right) f(\sigma) = \left(\frac{dv}{dt}\right)_{II} + p\sigma f(\sigma). \tag{1.5}$$

Since $p\sigma f(\sigma) > 0$ for $\sigma \neq 0$, then $\left(\frac{dv'}{dt}\right)_I$ turns out to be a positive definite function, q.e.d.

It should be mentioned that the additional Condition (1.2) does not limit the generality of the argument, since this is a necessary condition for System (I) to be asymptotically absolutely stable ([2], page 75).

If the parameters of System (II) satisfy the conditions previously established by Lur'e [1, 2] using Liapunov functions of the form (1.1), then the System (I) is asymptotically absolutely stable. It is proven below that, with a suitable choice of the arbitrary constant p, it becomes possible (except in certain special cases) to extend the region of absolute stability in parameter space. Let us carry out the analysis for the case h = 2.

2. Necessary Conditions for Asymptotic Absolute Stability

We consider the system

$$\dot{z}_\rho = \lambda_\rho z_\rho + f(\sigma) \quad (\rho = 1, 2),$$
$$\dot{\sigma} = \beta_1 z_1 + \beta_2 z_2 - r f(\sigma). \tag{2.1}$$

Replacing the nonlinear function $f(\sigma)$ by the linear one

$$f_*(\sigma) = h\sigma, \tag{2.2}$$

we obtain a linear system with the characteristic equation

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + h(r\lambda^2 + b_1\lambda + b_2) = 0, \tag{2.3}$$

where

$$a_1 = -(\lambda_1 + \lambda_2), \quad a_2 = \lambda_1\lambda_2,$$
$$b_1 = -(\lambda_1 + \lambda_2)r - (\beta_1 + \beta_2), \quad b_2 = \lambda_1\lambda_2 r + \beta_1\lambda_2 + \beta_2\lambda_1. \tag{2.4}$$

For the roots of Equation (2.3) to have negative real parts, it is necessary and sufficient that the following inequalities hold:

$$a_1 + rh > 0, \tag{2.5}$$
$$hb_2 > 0, \tag{2.6}$$
$$rh^2 b_1 + h(ra_2 + a_1 b_1 - b_2) + a_1 a_2 > 0. \tag{2.7}$$

We assume that $a_1 > 0$ and $a_2 > 0$. Inequalities (2.5) - (2.7) will be satisfied for all positive <u>h</u> if and only if for

$$r \geqslant 0 \tag{2.8}$$

and

$$b_2 > 0 \tag{2.9}$$

the following alternative holds:

$$b_1 > 0 \text{ and } ra_2 + a_1 b_1 - b_2 > -2\sqrt{ra_1 a_2 b_1}, \tag{2.10}$$

or

$$b_1 = 0 \text{ and } ra_2 \geqslant b_2.$$

The last inequalities (2.10) determine the presence either of real negative, or complex, roots of the trinomial expression (2.7) in <u>h</u>.

We rewrite Inequalities (2.10) in the following form:

$$b_1 > 0 \text{ and } \sqrt{b_2} < \sqrt{ra_2} + \sqrt{a_1 b_1},$$

or

$$b_1 = 0 \text{ and } b_2 \leqslant ra_2. \tag{2.11}$$

Inequalities (2.8), (2.9) and (2.11) are necessary for the trivial solution of System (2.1) to possess asymptotic absolute stability.

3. New Fomulation of the Lur'e Conditions

We consider the conditions for asymptotic absolute stability in the form [1, 2]:

$$r \geqslant 0, \tag{3.1}$$

$$\Gamma^2 = \frac{\beta_1}{\lambda_1} + \frac{\beta_2}{\lambda_2} + r > 0, \tag{3.2}$$

$$\text{Д}^2 = rf_1 - 2rf_2 + \beta_1\lambda_1 + \beta_2\lambda_2 \mp 2\varepsilon\sqrt{rf_2}\,\Gamma > 0 \tag{3.3}$$

(where $\epsilon = \pm 1$) and we express Conditions (3.2) and (3.3) as functions of a_1, a_2, b_1 and b_2 (Cf. paragraph 2). Making use of Relationships (2.4), we get

$$\Gamma^2 = \frac{b_2}{a_2} > 0, \tag{3.4}$$

$$-\left[b_2 - 2\varepsilon\sqrt{ra_2}\sqrt{b_2} + ra_2 - a_1b_1\right] > 0. \tag{3.5}$$

For Inequality (3.5) to be satisfied, it is necessary and sufficient that the roots of its left member, considered as a trinomial in $\sqrt{b_2}$, be real, and that the value of $\sqrt{b_2}$ lie between these roots:

$$b_1 > 0, \tag{3.6}$$

$$\varepsilon\sqrt{ra_2} - \sqrt{a_1b_1} < \sqrt{b_2} < \varepsilon\sqrt{ra_2} + \sqrt{a_1b_1}. \tag{3.7}$$

It is easily seen that the region corresponding to $\epsilon = -1$ is contained in that corresponding to $\epsilon = +1$, since $\sqrt{b_2} > 0$. Thus, the Lur'e conditions for $a_1 > 0$ and $a_2 > 0$ can be expressed in the following equivalent form:

$$\begin{gathered} r \geqslant 0,\ b_1 > 0,\ b_2 > 0, \\ \sqrt{ra_2} - \sqrt{a_1b_1} < \sqrt{b_2} < \sqrt{ra_2} + \sqrt{a_1b_1}. \end{gathered} \tag{3.8}$$

Comparing Inequalities (3.8) with the necessary Inequalities (2.8), (2.9) and (2.11), and excluding from consideration the boundary of the region of stability (for which $b_1 = 0$), we are led to the conclusion that for

$$a_1b_1 \geqslant ra_2 \tag{3.9}$$

the Lur'e conditions are not only sufficient, but also necessary. If (3.9) is not satisfied, the Lur'e conditions no longer coincide with the conditions for necessity. *

4. Extension of the Region of Absolute Stability

We now apply our lemma to System (2.1). Repeating our previous reasoning (Cf. paragraphs 2 and 3) for the case of a system of the form (II)

$$\begin{gathered} \dot{z}_\rho = \lambda_\rho z_\rho + f(\sigma) \quad (\rho = 1, 2), \\ \sigma = \beta_1\left(1 + \frac{p}{\lambda_1}\right)z_1 + \beta_2\left(1 + \frac{p}{\lambda_2}\right)z_2 - rf(\sigma), \end{gathered} \tag{4.1}$$

we obtain a characteristic equation of the form

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + h(r\lambda^2 + b_1\lambda + b_2') = 0. \tag{4.2}$$

---

* It was already remarked in [1] that the boundary of the region obtained by means of Lur'e's algorithm coincides with the envelope of the family (2.7).

Here,

$$b_1' = -r(\lambda_1+\lambda_2) - (\beta_1+\beta_2) - p\left(\frac{\beta_1}{\lambda_1}+\frac{\beta_2}{\lambda_2}\right) = b_1 + p\,\frac{ra_2-b_2}{a_2}, \tag{4.3}$$

$$b_2' = \lambda_1\lambda_2 r + \beta_1\lambda_2 + \beta_2\lambda_1 + p\left(\frac{\beta_1\lambda_2}{\lambda_1}+\frac{\beta_2\lambda_1}{\lambda_2}\right) = b_2 - p\,\frac{a_1b_2-a_2b_1}{a_2}, \tag{4.4}$$

where p is an arbitrary positive constant.

Using the Lur'e conditions in the form (3.8), we obtain

$$r_1 \geqslant 0,\ b_1' > 0,\ b_2' > 0, \tag{4.5}$$

$$\sqrt{ra_2} - \sqrt{a_1b_1'} < \sqrt{b_2'} < \sqrt{ra_2} + \sqrt{a_1b_1'}. \qquad (5.6$$

By virtue of the lemma, if Conditions (4.5), (4.6) and (3.4) are met, the latter condition being replaceable by

$$b_2 > 0, \tag{4.7}$$

since $a_2 > 0$, then System (2.1) is asymptotically absolutely stable.

5. Decomposition of the Region D

As before (Cf. paragraph 3), we exluded fromconsideration the boundary of the region of stability, $b_1 = 0$, and we consider, instead of the regions (2.8), (2.9) and 2.11), the region

$$r \geqslant 0,\quad b_1 > 0,\quad b_2 > 0, \qquad\qquad (5.1)$$
$$\sqrt{b_2} < \sqrt{ra_2} + \sqrt{a_1b_1}.$$

We decompose this region into three subregions:

subregion D_1, where

$$r \geqslant 0,\ b_1 > 0,\ b_2 > 0,$$
$$\sqrt{ra_2} - \sqrt{a_1b_1} < \sqrt{b_2} < \sqrt{ra_2} + \sqrt{a_1b_1}; \qquad (5.2)$$

subregion D_2, where

$$r \geqslant 0,\ b_1 > 0,\ b_2 > 0, \qquad\qquad (5.3)$$
$$\sqrt{b_2} \leqslant \sqrt{ra_2} - \sqrt{a_1b_1},\qquad a_1b_2 - a_2b_1 \leqslant 0;$$

and subregion D_3, where

$$r \geqslant 0,\ b_1 > 0,\ b_2 > 0, \qquad\qquad (5.4)$$
$$\sqrt{b_2} \leqslant \sqrt{ra_2} - \sqrt{a_1b_1},\qquad a_1b_2 - a_2b_1 > 0.$$

It is easy to verify that the sum of the regions D_1, D_2 and D_3 equals

$$D = D_1 \cup D_2 \cup D_3. \tag{5.5}$$

In region D_1 System (2.1) is asymptotically absolutely stable, by virtue of the Lur'e Conditions (3.8). Regions D_2 and D_3 will be investigated below.

6. Absolute Stability of System (2.1) in Region D_2

Indeed, in region D_2, b_1' and b_2' [Cf. (4.3) and (4.4)] are positive for all positive p and, moreover, they are

both monotone nondecreasing for increasing p. For $p = 0$, Inequality (4.6) does not hold in region D_2. For increasing p, the leftmost member decreses monotonically to $-\infty$, the quantity $\sqrt{b_2'}$ remains monotone nondecreasing, and the rightmost quantity increases monotonically. Therefore, there is a positive value of p for which Inequality (4.6) holds. Since (4.7) also holds in region D_2, all the conditions (4.5), (4.6) and (4.7) are satisfied in region D_2 and, consequently, System (2.1) is asymptotically absolutely stable in region D_2.

7. Absolute Stability of System (2.1) in Region D_3

In the region, b_1' is a monotonic increasing, and b_2' as montonic decreasing function of p. Therefore, it is necessary that the following inequality be satisfied:

$$p < \frac{a_2 b_2}{a_1 b_2 - a_2 b_1}. \tag{7.1}$$

Let us rid ourselves of the last constraint.

If the following inequality is satisfied:

$$\left(\sqrt{ra_2} = \sqrt{a_1 b_1'}\right)^2 < b_2' < \left(\sqrt{ra_2} + \sqrt{a_1 b_1'}\right)^2, \tag{7.2}$$

the Inequalities (4.6) and (7.1) are also satisfied, since from (7.2) it follows that $b_2' > 0$.

If the inequalities $r \geq 0$, $b_1' > 0$, $b_2 > 0$ and (7.2) are all satisfied, then the conditions established above (Cf. paragraph 4) are also satisfied, and therefore the trivial solution of System (2.1) possesses asymptotic absolute stability.

Condition (7.2), having the form:

$$M^2 < b_2' < N^2, \tag{7.3}$$

is equivalent to the condition

$$(b_2')^2 - (M^2 + N^2) b_2' + M^2 N^2 < 0. \tag{7.4}$$

Making use of Relationships (4.3) and (4.4) and omitting the calculations, we obtain the inequality

$$Ap^2 + Bp + C < 0, \tag{7.5}$$

where

$$A = (ra_1 - b_1)^2, \tag{7.6}$$

$$\frac{1}{2} B = a_1 b_1 (ra_1 - b_1) - (ra_1 + b_1)(ra_2 - b_2), \tag{7.7}$$

$$C = (ra_2 - b_2)^2 + a_1 b_1 (a_1 b_1 - 2ra_2 - 2b_2). \tag{7.8}$$

The trinomial on the left of Inequality (7.5) always has real roots in region D_3, since

$$B^2 - 4AC = 16 r a_1 b_1 \left[(ra_2 - b_2)^2 + (ra_1 - b_1)(a_1 b_2 - a_2 b_1)\right] > 0. \tag{7.9}$$

In fact, in region D_3 [Cf. (5.4)] we have $b_2 < ra_2$ and $b_2 > \frac{a_2 b_1}{a_1}$, and consequently,

$$ra_1 - b_1 > 0. \tag{7.10}$$

It remains to prove that in region D_3 there exists a positive value of p for which Inequality (7.5) holds.

We noticed that, for $b_1 = 0$, the trinomial on the left of Inequality (7.5) has the multiple root

$$p_{1,2} = \frac{ra_2 - b_2}{ra_1} > 0. \tag{7.11}$$

For increasing b_1, the roots of the trinomial become distinct, although remaining real [Cf. (7.9)]. Due to the continuous dependence of the roots on the coefficients of the trinomial, the roots will remain within the region $p > 0$ for sufficiently small b_1. If one of the roots of the trinomial in (7.5) becomes zero, then C = 0. The least value of b_1 [Cf. (7.8)] for which this equality can hold will be

$$b_1 = \frac{1}{a_1}(\sqrt{ra_2} - \sqrt{b_2})^2, \tag{7.12}$$

but this is also the greatest value that b_1 can assume in the region D_3 [Cf. (5.4)]. Consequently, in the region D_3 there is a positive value of p for which (7.5) is satisfied, and therefore System (2.1) is asymptotically absolutely stable. Since System (2.1) is asymptotically absolutely stable in regions D_1, D_2 and D_3, then by virtue of (5.5) we can formulate our conslusions, as detailed below.

8. Necessary and Sufficient Conditions for Asymptotic Absolute Stability

If it is assumed that $a_1 > 0$ and $a_2 > 0$, and if we exclude from consideration the boundary of the region of stability ($b_1 = 0$), then the necessary and sufficient conditions for asymptotic absolute stability are:

$$r \geqslant 0,\ b_1 > 0,\ b_2 > 0,\ \sqrt{b_2} < \sqrt{ra_2} + \sqrt{a_1 b_1}, \tag{8.1}$$

where b_1, b_2, a_1 and a_2 are the coefficients of the characteristic Equation (2.3) obtained from System (2.1) after linearization by means of the function $f_*(\sigma) = h\sigma$.

Thus, it has been established that for three differential equations the lemma, proved in this article, allows one to obtain the region of asymptotic absolute convergence coinciding (with the exception of some special cases which require individual investigation) with the largest region of asymptotic absolute convergence which may be obtained in parameter space.

The author desires to thank the investigators at the Institute Matematiki Akademii Nauk Rumynskoi Narodnoi Respubliki, working in the domain of differential equations, for their kind attention and valued comments.

LITERATURE CITED

[1] A. M. Letov, Stability of Nonlinear Control Systems* (Gostekhizdat, 1955).

[2] A. I. Lur'e, Some Nonlinear Problems in the Theory of Automatic Control* (Gostekhizdat, 1951).

Received August 20, 1957

* In Russian.

OPTIMAL TRANSIENTS IN AN AUTOMATIC CONTROL SYSTEM HAVING A BOUNDED REGULATOR UNIT

E. K. Krug and O. M. Minina

(Moscow)

The forms of the optimum transients in automatic control systems (ACS) are determined for those with objects of various response characteristics (including delays), assuming the control valve to be position bounded. It is shown that there are difficulties in using continuous-action regulators to produce optimal transients, because the nonlinear converters to the regulators must have responses determined by the magnitude and site of the perturbations, and by the initial values of the bounded co-ordinates. Stepwise operated regulators are proposed for use in obtaining optimal transients.

In designing ACS one has to consider how to ensure the minimum duration of transients produced by perturbations. The condition that the controlled quantity differ minimally from its set value is applied.

To solve this problem one must analyze the behavior when the regulator is position bounded and various types of perturbation act, and produce transients of minimum duration.

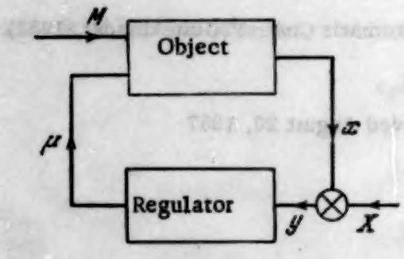


Fig. 1. Structural diagram of an ACS.

Such transients will in future be termed optimal, the transient duration being taken as the time from when the perturbation is applied to when the controlled quantity shows precisely zero deviation. The latter situation can, as we know, occur in nonlinear systems. We must also determine the regulator response which provides optimal transients, and find ways of realizing it.

Here we consider an ACS of block diagram as in Fig. 1. Here two types of perturbation act, one on the load (M) and the other on the input (X). The optimal transient responses will be found for stepwise perturbations to load and input for objects which are described by first-order differential equations with delays, and also for objects specified by several first-order linkages. We suppose the regulator to have two extreme (bounded) positions. A fast regulator is assumed, i.e., one with time constants incomparably smaller than those of the object.

1. Derivation of the Optimal Transients in an Automatic Control System Having a Bounded Regulator Unit.

In [1] we considered a linear ACS consisting of an object with a lag and a fast-acting regulator. It was shown that there was a limit to the transient duration even if the gain were infinite when there was a lag. The duration of the transient produced by a step perturbation to the load was 2τ, or τ if to the input, τ being the lag in the object (Fig. 2). The maximum deviation from the set value in the output (y_{max}) when the perturbation is to the load can be determined if the response of the object and the magnitude of the perturbation are known. Thus for a first-order non-self-regulating object with a lag the maximum deviation (Fig. 2) can be found from

$$y_{\max} = \Delta M \frac{\tau}{T}, \tag{1}$$

while for a first-order self-regulating object with a lag we have

$$y_{\max} = \Delta M k (1 - e^{-\frac{\tau}{T}}), \tag{2}$$

where $\Delta M = M - M_0$, M_0 specifying the state before the perturbation, T being the time-constant of the object and k the gain in the object.

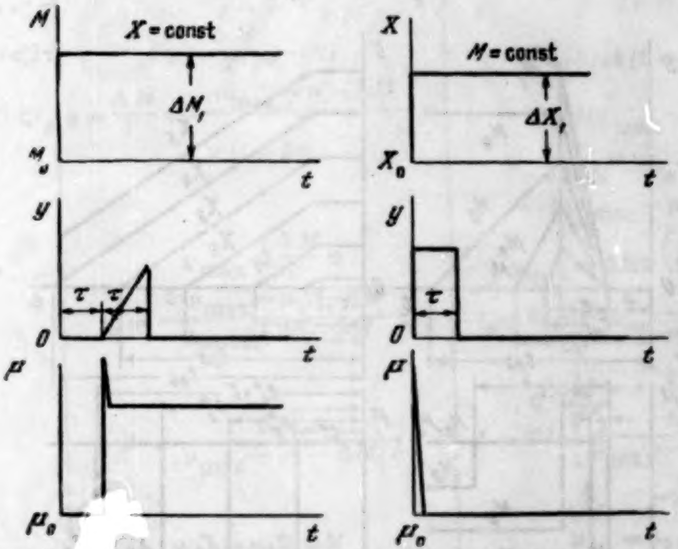


Fig. 2. Optimal transient response in an ACS of infinite gain.

A regulator which is to provide the optimum transient response to such perturbations must, when the output deviates from the set value ($y \neq 0$), apply an infinite control signal and instantly set the controller in the position corresponding to the new equilibrium state.

The fact that coordinates and their derivatives are always bounded in real systems is neglected in [1]. Even if the regulator is very quick-acting and shows no coordinate bounding we cannot obtain the transients shown in Fig. 2 because the controller (throttle valve, slide valve, etc.) is position-bounded. This corresponds to the μ coordinate (Fig. 1) being bounded.

We shall derive the optimum transient response for the system shown schematically in Fig. 1 for a position-bounded controller. We shall assume the system in a steady state and that step perturbations to load and input are applied. It has been shown [2, 3] that the system will be of quickest response if the bounded coordinates have their maximum values during the transient. Hence in an ACS having an object with a first-order delay, the regulator must set the controller at the appropriate maximum displacement $\mu_{\max}$ when a perturbation acts at $t = 0$ to the input or at $t = \tau$ to the load, retain this setting for a definite time, and then set it at the position μ_{ss}, corresponding to the steady state.

In an ACS with such a regulator the maximum deviation $y_{\max}$, the transient duration t_p and the time the controller is held at the extreme position t^* are related by the relationships given in Table 1 (μ_0 being the controller position before the perturbation).

Table 1 is easily derived if we consider the transient in sections divided according to the controller position and assume the controller restored to the equilibrium position at a time τ before the end of the transient.

Fig. 3 shows the optimal transients for various step perturabations to the load for a nonself-balancing object with a first-order delay $T/\tau = 6$ (Fig. 3a), or for perturbations to the input (Fig. 3b).

The controller position bounding increases the duration of the optimal transient over that for an unbounded system (Fig. 2), but the maximum deviation is unaltered.

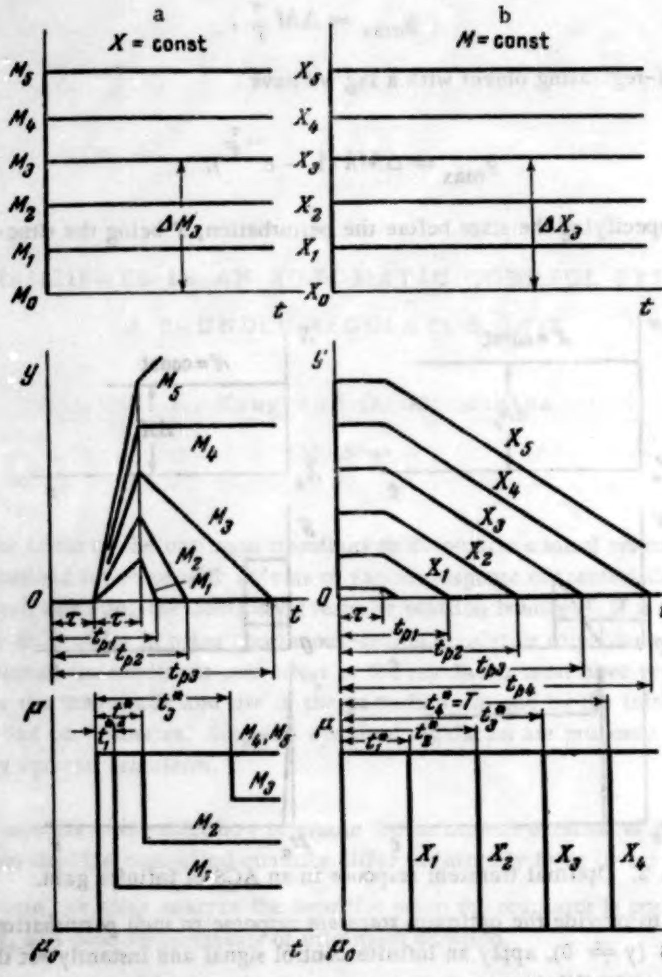


Fig. 3. Optimal transients in an ACS with a position-bounded controller.

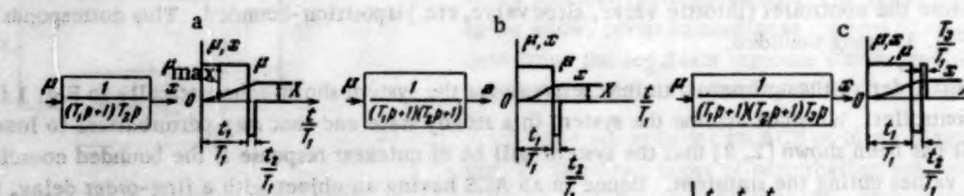


Fig. 4. Extreme position dwell time for controller vs perturbation;
a) for a second-order object without self-regulation, b) for a second-order object with self-regulation , c) for a third-order object without self-regulation.

The transients shown in Fig. 3 and Table 1 relate to alternate perturbations to a system initially at equilibrium ($y = 0$, $dy/dt = 0$, $\mu = \mu_0$).

TABLE 1

Parameter	Object with a first-order delay without self-regulation	Object with a first-order delay with self-regulation
	Step perturbation to load	
y for $0<t<\tau$	$y=0$	$y=0$
y for $\tau<t<2\tau$	$y=\frac{\Delta M}{T}(t-\tau)$	$y=\Delta Mk\left(1-e^{-\frac{t-\tau}{T}}\right)$
y for $2\tau<t<t_p$	$y=\frac{\Delta M}{T}\tau-\frac{\mu_{max}-\mu-\Delta M}{T}\times(t-2\tau)$	$y=k\left(\mu_{max}-\mu_0-\Delta Me^{-\frac{\tau}{T}}\right)e^{-\frac{t-2\tau}{T}}-k(\mu_{max}-\mu_0-\Delta M)$
y_{max}	$y_{max}=\frac{\Delta M}{T}\tau$	$y_{max}=\Delta Mk\left(1-e^{-\frac{\tau}{T}}\right)$
t_p	$t_p=\frac{2(\mu_{max}-\mu_0)-\Delta M}{\mu_{max}-\mu_0-\Delta M}\tau$	$t_p=2\tau+T\ln\left[\frac{\Delta M}{\mu_{max}-\mu_0-\Delta\mu}\times\left(1-e^{-\frac{\tau}{T}}\right)+1\right]$
t^*	$t^*=\frac{\Delta M}{\mu_{max}-\mu_0-\Delta M}\tau$	$t^*=T\ln\left[\frac{\Delta M}{\mu_{max}-\mu_0-\Delta M}\times\left(1-e^{-\frac{\tau}{T}}\right)+1\right]$
μ_{ss}	$\mu_{ss}=\mu_0+\Delta M$	$\mu_{ss}=\mu_0+\Delta M$
	Step perturbation to input	
y for $0<t<\tau$	$y=\Delta X$	$y=\Delta X$
y for $\tau<t<t_p$	$y=\Delta X-\frac{\mu_{max}-\mu_0}{T}(t-\tau)$	$y=\Delta X-k(\mu_{max}-\mu_0)\left(1-e^{\frac{\tau-t}{T}}\right)$
y_{max}	$y_{max}=\Delta X$	
t_p	$t_p=\tau+\frac{\Delta XT}{\mu_{max}-\mu_0}$	$t_p=\tau+T\ln\frac{\mu_{max}-\mu_0}{\mu_{max}-\mu_0-\frac{\Delta X}{k}}$
t^*	$t^*=\frac{\Delta XT}{\mu_{max}-\mu_0}$	$t^*=T\ln\frac{\mu_{max}-\mu_0}{\mu_{max}-\mu_0-\frac{\Delta X}{k}}$
μ_{ss}	$\mu_{ss}=\mu_0$	$\mu_{ss}=\mu_0+\frac{\Delta X}{k}$

If perturbations are applied to load and input the relationships of t^*, t_p and y_{max} will be different. E. g., if both act together and are steps the extreme position dwell time can be found from

$$t^*=\frac{\Delta XT}{\mu_{max}-\mu_0-\Delta M}. \tag{3}$$

If the load step occurs at a time τ before the input step we should use

$$t^*=\frac{\Delta XT+\Delta M\tau}{\mu_{max}-\mu_0-\Delta M}. \tag{4}$$

From the above we may draw up the requirements for a regulator which will provide the optimum transient in a position-bounded controller system containing a first-order object with a delay.

1. When the controlled quantity deviates from its set value the regulator must instantly set the controller in one of the limiting positions, if both perturbations act together.

2. The controller must be held at the extreme position for a definite time t^*, depending on the object parameters and the type of perturbation.

3. The controller must then be set at μ_{ss}, corresponding to the new steady state in the system.

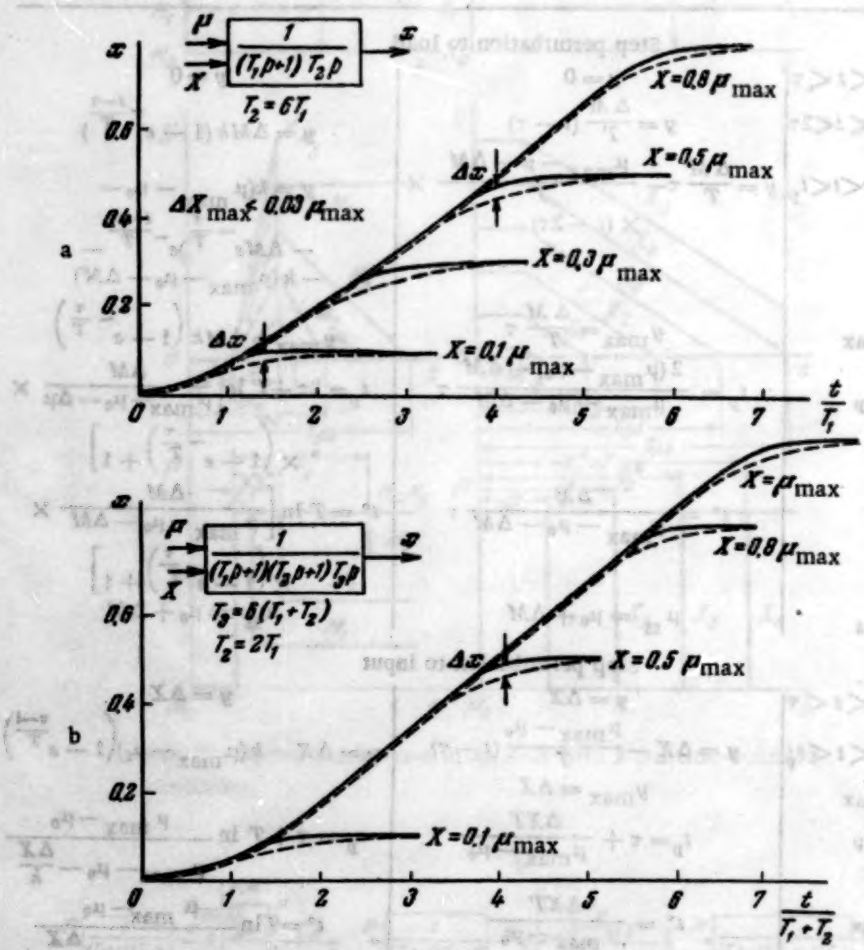


Fig. 5. Transients with various types of controller reversals. Full lines: a) two, b) three reversals; dashed lines, one reversal ($t_1 = t^* = XT/\mu_{max}$).

A regulator which fulfills these functions can only produce optimal transients in an ACS containing a first-order object with a delay. We will determine what transients we will get with such a controller in an ACS with a second-order object, and how they differ from the optimal ones.

The optimal transients in an ACS containing an n-th order object and a position-bounded controller can be obtained by switching the controller n times from one limiting position to the other. Fig. 4 gives calculated curves for the relation of controller extreme position dwell time to input signal for various objects. It is clear that if one of the time-constants in the object is greater than the others the transient duration ($t_1 + t_2 + t_3 + \ldots$) is mainly determined by the dwell time for the first position (t_1).

Fig. 5 shows the optimal transients in systems with second- and third-order objects, using numerical examples. Transients for the same ACS with one controller reversal are also shown, the controller dwell time being found from the expression for t^* (Table 1), by replacing T by the largest time-constant, and τ by the sum of the others (noting that $t^* < t_1$).

The examples of Fig. 5 enable us to derive the deviations from the optimal transients in the latter (one reversal) instance, for high-order objects. The examples given in [4] indicate that the actual transients with one reversal are close to optimal for their high-order objects. Hence we may assume that the regulator problem for such objects is very similar to that for first-order objects.

The regulator problem may be solved in two ways.

1. By inserting nonlinear converters in a continuous-acting regulator [isodrome (proportional + floating), derivative, etc.]. This regulator must ensure that the controller is set in definite positions at set times and provide a stable steady state.

2. By producing a discontinuous-acting regulator with the same functions.

2. Inserting Nonlinear Converters in a Continuous-Acting Regulator to Obtain Optimal Transients.

As it is difficult to produce signals proportional to the second, third, etc., derivatives of the controlled coordinate in real apparatus, the simplest continuous-acting regulators have been widely used in practice (isodrome, proportional, first-derivative). Such systems are analyzed and approved from linear studies [5, 6], the nonlinearities occurring in real systems being neglected.

It has been shown [7] that controller position bounding increases the transient duration and deviation greatly with large perturbations.

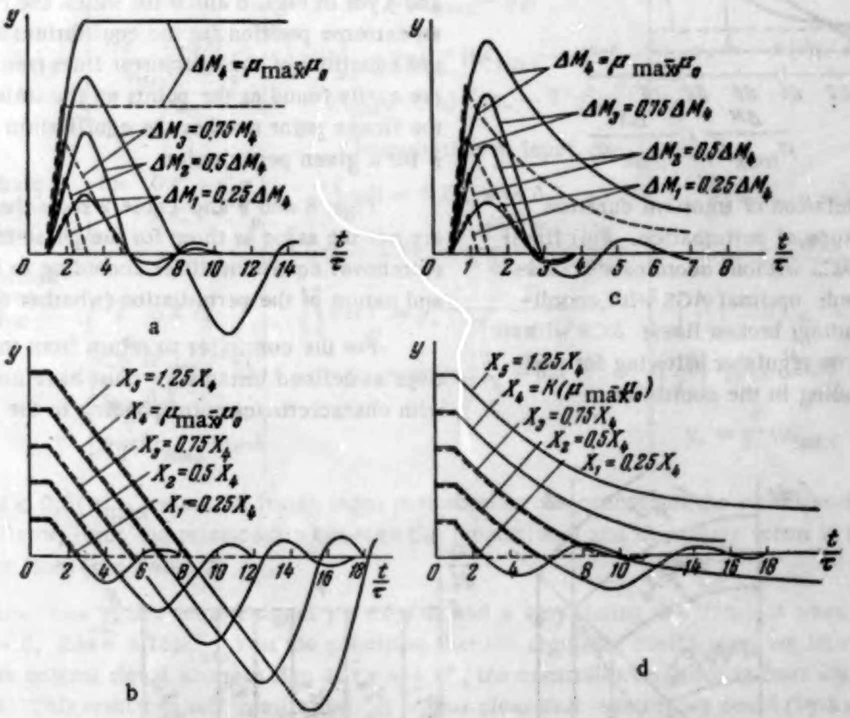


Fig. 6. Transients for an ACS with a first-order object and a delay:
a) without self-balancing, perturbation to load, b) the same, perturbation to input, c) with self-balancing perturbation to load, d) the same, perturbation to input.

Fig. 6. shows the optimal transients for an ACS with a first-order object with a delay (dashed) plus those in an ACS with an isodrome regulator of parameters selected in accordance with Ziegler and Nichols' rules [5] (full lines). Fig. 7 shows how t_p varies with the size of the perturbation for an ACS with a first-order object (non-self-balancing) and a delay.* It is clear that the transient duration is reduced by a large factor if the regu-

* t_p for an ACS with an isodrome regulator is considered as the time from when the perturbation is applied to when $y = 0.01\ X_4$.

lator provides optimal transients.

We may wonder whether the transients may be brought near to optimal by using nonlinear converters in continuous-acting regulators.

The nonlinear relationships required for this purpose can be found by considering the behavior during optimal transients in the (dy/ dt, y) and (y, ∫ydt) planes. This is possible because the controller has definite fixed positions (μ_{max}, μ_0, μ_{ss}) in defined time intervals.

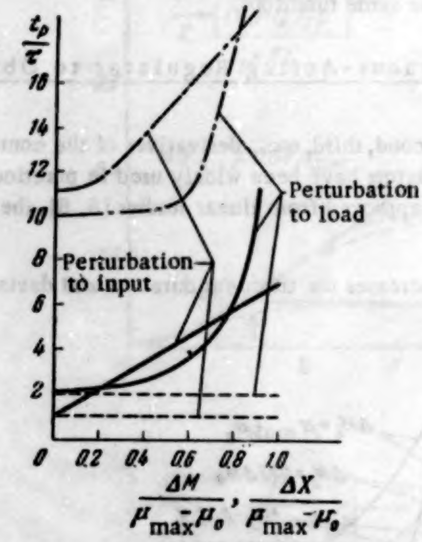


Fig. 7. Relation of transient duration to magnitude of perturbation. Full lines: optimal ACS without coordinate bounding; dashed: optimal ACS with coordinate bounding; broken lines: ACS with an isodrome regulator allowing for position-bounding in the controller.

Table 2 summarizes the equations describing the phase behavior in these planes for the above objects in relation to magnitude of perturabation. Figs. 8 and 9 give the phase trajectories for various perturbations. The initial conditions for input perturbations are taken as

$$y=\Delta X,\ dy/dt=0,\ \mu=\mu_0.$$

For load perturbations (Fig. 9) the initial conditions are

$$y=0,\ dy/dt=\Delta M/T,\ \mu=\mu_0.$$

The relations between dy/ dt and y or between y and ∫ydt in Figs. 8 and 9 for which the controller exchanges an extreme position for the equilibrium one are termed the equations of the switchover lines (see Table 2). These are easily found as the points on the trajectories from which the image point reaches the equilibrium position in a time τ for a given perturbation.

Figs. 8 and 9 and Table 2 show that these equations are not the same as those for the phase trajectories. The switchover equations differ according to the magnitude and nature of the perturbation (whether to load or input).

For the controller to return from the extreme positions at defined instants we must have nonlinear converters with characteristics corresponding to the switchover lines.

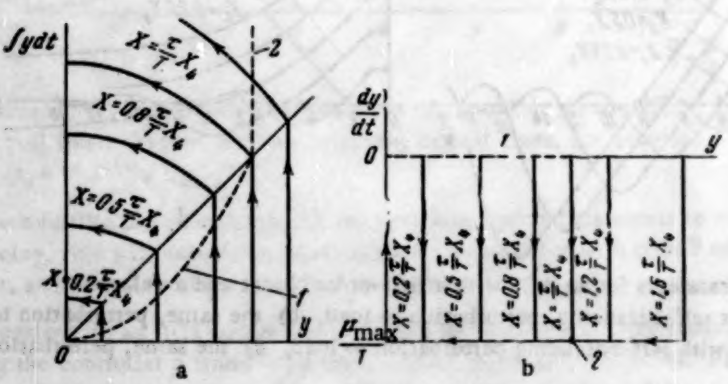


Fig. 8. Phase pictures for input perturbations.

The dashed lines are for switchovers when $X < \frac{\tau}{T} X_4$ (curve 1).

The case $X > \frac{\tau}{T} X_4$ corresponds to curve 2 in Fig. 8a and point 2 in in 8b. In Figs. 8a and b $X_0 = 0$, $X_4 = \mu_{max} - \mu_0$

If, for instance, we use a first-derivative regulator with control law

$$\mu = k \ (y + \tau dy / dt), \tag{5}$$

and if the gain k is large enough (as soon as $y \neq 0$ the controller moves to the extreme position), then for a load perturabation $\Delta M > 0.5\,(\mu_{max} - \mu_0)$ the regulator will hold the controller fully over for a time $t = t^*$ such as to give the optimal transient.

TABLE 2

Parts of phase picture		Equations	
		in the plane $\int ydt;\, y$	in the plane $\frac{dy}{dt};\, y$
		Perturbation to load	
Phase trajectories	for $\tau < t < 2\tau$	$\int ydt = 0,5 \frac{T}{\Delta M} y^2$	$\frac{dy}{dt} = \frac{\Delta M}{T}$
	» $2\tau < t < t_p$	$\int ydt = \frac{\Delta M \tau^2 (\mu_{max} - \mu_0) - T^2 y^2}{2T(\mu_{max} - \mu_0 - \Delta M)}$	$\frac{dy}{dt} = \frac{\Delta M - \mu_{max} + \mu_0}{T}$
Switchover line	» $-\Delta M < \frac{\mu_{max} - \mu_0}{2}$	$\left[\int ydt\right]_* = \frac{Ty^2 + Ty_*^{3/2}\sqrt{y_* + 4\frac{\tau}{T}(\mu_{max} - \mu_0)}}{4(\mu_{max} - \mu_0)}$	$y_* = \frac{[dy/dt]_*^2 \tau}{\frac{\mu_{max} - \mu_0}{T} - [dy/dt]_*}$
	for $-\Delta M > \frac{\mu_{max} - \mu_0}{2}$	$\left[\int ydt\right]_* = \frac{\tau}{2T^2 y_*}[(\mu_{max} - \mu_0)^2\tau^2 - (\mu_{max} - \mu_0)T\tau y_* - T^2 y_*^2]$	$y_* = -\tau\,[dy/dt]_*$
		Perturbation to input	
Phase trajectories	for $0 < t < \tau$	$\int ydt = \Delta X t,\ y = \Delta X$	$\frac{dy}{dt} = 0,\ y = \Delta X$
	» $\tau < t < t_p$	$\int ydt = \Delta X \tau + \frac{T(\Delta X - y^2)}{2(\mu_{max} - \mu_0)}$	$\frac{dy}{dt} = -\frac{\mu_{max} - \mu_0}{T}$
Switchover line	» $\Delta X < \frac{\tau}{T}(\mu_{max} - \mu_0)$	$\left[\int ydt\right]_* = y_*^2 \frac{T}{\mu_{max} - \mu_0}$	$\left[\frac{dy}{dt}\right]_* - 0,\ y_* = \Delta X$
	for $\Delta X > \frac{\tau}{T}(\mu_{max} - \mu_0)$	$y_* = \frac{\tau}{T}(\mu_{max} - \mu_0)$	$\left[\frac{dy}{dt}\right]_* = -\frac{\mu_{max} - \mu_0}{T},\ y_* = \frac{\tau}{T}(\mu_{max} - \mu_0)$

When $\Delta M < 0.5(\mu_{max} - \mu_0)$, or for an input perturbation, we cannot get the corresponding optimal transient with such a regulator, since the relationship between the proportional and derivative terms in (5) will not correspond to the switchover lines (see Table 2).

Fig. 10 shows how y, the control signal $y + \tau\, dy/dt$ and μ vary during the transient when a step load perturbation occurs ($\mu_0 = 0$, $\Delta M = 0.75\mu_{max}$) on the condition that the regulator coefficients are related as in (5). Fig. 10 shows that the control signal changes sign at $t = \tau + t^*$, the controller transferring from $+\mu_{max}$ to $-\mu_{max}$, and not to $\mu_{ss} = \Delta M$. This results in self-oscillation. It is thus clear that even if we could (by knowing the type and size of perturbation and initial controller position) reset the controller from the extreme position at a given moment, we cannot thereby get it back to the steady state; this results in self-oscillation.

If the switchover functions are generated by using nonlinear converters in the isodrome (proportional + floating) regulators the resulting system is structurally unstable. In fact, if the control signal, consisting of proportional and integral terms, is to be zero at some set time the integral term must have the reverse sign.

These peculiarities, particularly the fact that the nonlinear converters are not single-valued, which is related to the differing initial conditions, and the instability in the equilibrium state, do not only occur in systems with delays.

Let us consider a system with a second-order object and a position-bounded controller. To obtain optimal transients here we must have two times (t_1 and t_2, Fig. 6b) when the controller is reset to extreme positions.

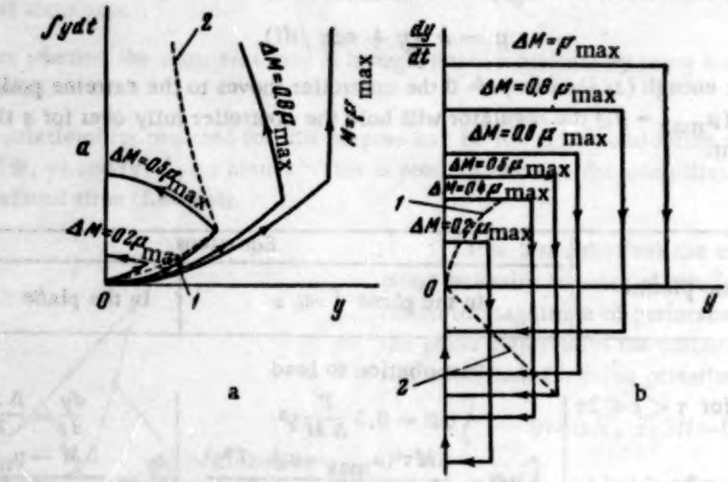


Fig. 9. Phase pictures for load perturbations.
The dashed lines are for switchovers when $\Delta M < 0.5\,\mu_{max}$ (curve 1) and when $\Delta M > 0.5\mu_{max}$ (curve 2).

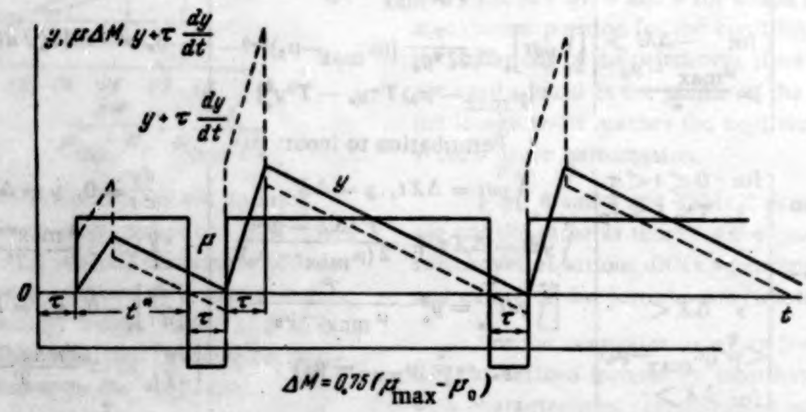


Fig. 10. Transients in an ACS with non-self-regulating first-order object with a delay.

Fig. 11 shows the phase trajectories corresponding to the optimal transients in response to input perturbations for two initial states ($\mu_0 = 0$ and $\mu_0 = 0.75\,\mu_{max}$) in dy/dt and y coordinates. The switchover line (dashed in Fig. 11) is not single-valued, and also depends on the perturbation and initial state.

If the controller here switches over at given times but does not reach the equilibrium position at $t = t_1 + t_2$ we shall again have self-oscillation.

To produced optimal transients in an ACS the continuous-acting regulators plus nonlinear converters must contain elements which effect the switchover functions as appropriate to the magnitude, type and site of the perturbation and to the controller position. They must also contain elements which enable one to set the controller in a position corresponding to the perturbations applied. This rules out the regulator operating continuously during the transient.

However, this does not imply that we cannot improve the response of an ACS subject to perturbations of a single species by using nonlinear links.

3. ACS with a Discrete-Action Regulator

Here we describe the operation of a discrete-action regulator with which we can obtain optimal transients in an ACS containing a first-order object with a delay, and also consider examples of the transients resulting from various perturbations.

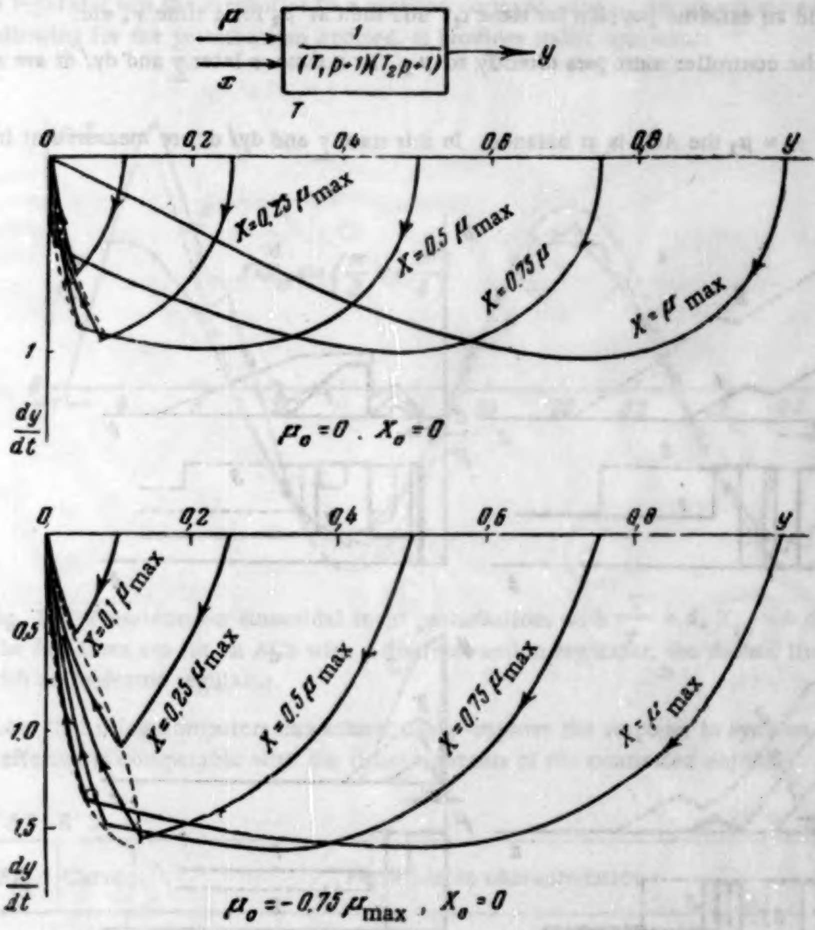


Fig. 11. Phase trajectories for an optimal ACS with a second-order object.

The conclusions on the optimal transient in response to load perturbations are drawn on the assumption that the magnitudes of these are not measured directly; but the expressions of Section 1 ((3) - (4), Table 1) contain the load perturbation.

If the load perturbation can be measured we can design a regulator which can balance out the perturbation, without interference from the lag in the object. The transient duration will then be determined only by the regulator lag, and systems containing objects with very different parameters and quick-acting regulators will only show transients in response to input perturbations.

In an ACS of object response described by linear equations we can evaluate the load changes (if the object parameters are known) from the way the controlled coordinate and its derivatives change. For first-order objects with delays the derivative of the deviation specifies the load change in the time preceding the derivative determination ($\Delta M = T dy/dt$). In such cases, by measuring the controlled coordinate and its derivatives we can design a discrete-action regulator having the functions stated in Section 1.

In an ACS of this type t^* and μ_{ss} can be found from suitable formulas using a computing unit.

The computer must be supplied with $\underline{y}$ and dy/dt at defined times. These times must not correspond to extreme positions in the controller. With objects having delays it is advantageous to measure $\underline{y}$ and dy/dt at a time τ after the controller has been adjusted from an extreme to a calculated position.

The requirements as to the functions of a discrete-action regulator are thus as follows: if the ACS was at equilibrium ($y = 0$, $dy/dt = 0$, $\mu = \mu_0$) and the output deviates because of some perturbations, then the regulator, having measured some y_1 and dy_1/dt, must set the controller in an extreme position, hold it there for a time t_1^* and then set it at the new position μ_1. At time τ after setting the controller to μ_1 the regulator must measure fresh values y_2 and dy_2/dt (again test the system), and from these calculate t_2^*; if t_2^* differs from zero

the controller is set in an extreme position for time t_2^* and then at μ_2 for a time τ, etc.

If t_2^* is zero the controller must pass directly to μ_2. At a time τ later y and dy/dt are again determined, etc.

If $t_2^* = 0$ and $\mu_2 = \mu_1$ the ACS is at balance. In this state y and dy/dt are measured at intervals of τ.

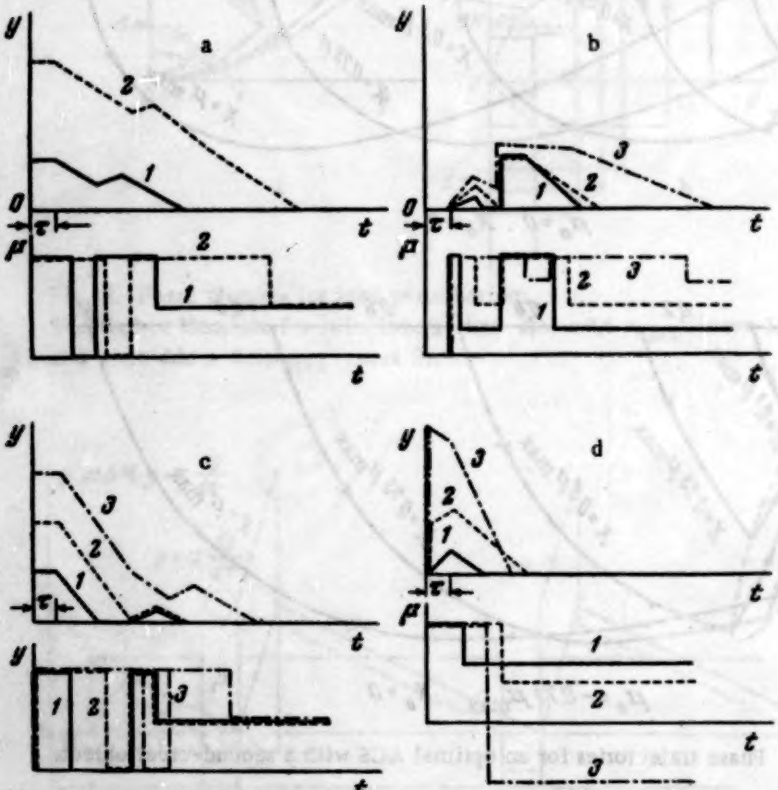


Fig. 12. Transients in an ACS with a discrete-action regulator.

If any additional perturbation occurs between tests (perturbations applied subject to arbitrary initial conditions) the system will react next time y and dy/dt are tested.

In this ACS the discrete-action regulator will sample y and dy/dt at intervals of $\tau + t^*$. If the perturbations are small the intervals between tests will be close to τ.

An ACS with a first-order non-self-balancing object with a delay has relations which define t^* and μ_{ss} of the form

$$t^* = \frac{(y + \tau\, dy/dt)\, T}{\mu_{max} - \mu_i - T\, dy/dt}, \tag{6}$$

$$\mu_{ss} = \mu_{i+1} = \mu_i + T\, dy/dt \quad (i = 0, 1, 2, 3, \ldots). \tag{7}$$

Figs. 12 and 13 show the transients in an ACS with the above discrete-action regulator for step perturbations of different sizes simultaneously applied (Fig. 12a), or applied at different instants (Fig. 12, b and c), for uniformly changing perturbations to the input (Fig. 12d), and for sinusoidal perturbations to the input (Fig. 13).

Table 3 gives the conditions on which Fig. 12 was drawn up.

Thus a regulator of this type can provide optimal transients to step perturbations to load and input and to uniformly changing input perturbations.

We would expect such a regulator to give near-optimal transients with other slowly varying perturbations, since the interval between tests depends on the deviation in the output and on its derivative.

Since this regulator sets the controller in a position corresponding to the steady state for a time τ between samples while allowing for the perturbations applied, it provides stable operation.

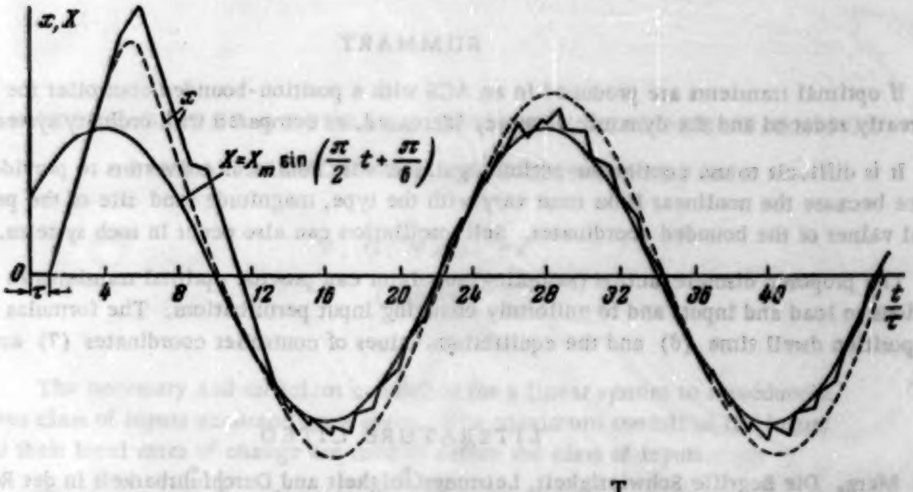


Fig. 13. Transients for sinusoidal input perturbations with $\frac{T}{\tau} = 6$, $X_m = 0.4(\mu_{max} - \mu_0)$. The full lines are for an ACS with a discrete-action regulator, the dashed lines for one with an isodrome regulator.

We also note that using computers can substantially improve the response in systems in which the time taken to displace the effector is comparable with the time-constants of the controlled object.

TABLE 3

Fig.	Curve	Perturbation characteristics
12, a	1,2	$M = M_0$ and $X = X_0$ for $t<0$, $M = M_0 + \Delta M$ and $X = X_0 + \Delta X$ for $t>0$
	1	$\Delta M = 0.5\ (\mu_{max} - \mu_0)$, $\Delta X = 0.25\ (\mu_{max} - \mu_0)$
	2	$\Delta M = 0.5\ (\mu_{max} - \mu_0)$, $\Delta X = 0.75\ (\mu_{max} - \mu_0)$
12, b	1,2,3	$M = M_0$ for $t<0$, $M = M_0 + \Delta M$ for $t<0$, $X = X_0$ for $t<3\tau$, $X = X_0 + \Delta X$ for $t>3\tau$
	1	$\Delta X = 0.25\ (\mu_{max} - \mu_0)$, $\Delta M = 0.25\ (\mu_{max} - \mu_0)$
	2	$\Delta X = 0.25\ (\mu_{max} - \mu_0)$, $\Delta M = 0.5\ (\mu_{max} - \mu_0)$
	3	$\Delta X = 0.25\ (\mu_{max} - \mu_0)$, $\Delta M = 0.75\ (\mu_{max} - \mu_0)$
12, c	1,2,3	$X = X_0$ for $t<0$, $X = X_0 + \Delta X$ for $t>0$, $M = M_0$ for $t<3\tau$, $M = M_0 + \Delta M$ for $t>3\tau$
	1	$\Delta M = 0.5\ (\mu_{max} - \mu_0)$, $\Delta X = 0.25\ (\mu_{max} - \mu_0)$
	2	$\Delta M = 0.5\ (\mu_{max} - \mu_0)$, $\Delta X = 0.5\ (\mu_{max} - \mu_0)$
	3	$\Delta M = 0.5\ (\mu_{max} - \mu_0)$, $\Delta X = 0.75\ (\mu_{max} - \mu_0)$
12, d	1,2,3	$M = M_0$ — const, $X = X_0$ for $t<0$, $X = X_0 + \Delta X + k_x t$ for $t>0$
	1	$\Delta X = 0$, $k_x = 0{,}1\ \frac{\mu_{max} - \mu_0}{\tau}$
	2	$\Delta X = 0.25\ (\mu_{max} - \mu_0)$, $k_x = 0.07\ \frac{\mu_{max} - \mu_0}{\tau}$
	3	$\Delta X = 0.75\ (\mu_{max} - \mu_0)$, $k_x = 0.1\ \frac{\mu_{max} - \mu_0}{\tau}$

The relations defining t^*, which allow for speed bounding in the effector, are easily derived for step perturbations to load and input.

Relations giving t^* for first-order self-regulating objects with delays, and for the objects of higher order considered at the end of Section 1, are also easily derived. In the latter case the expressions for t^* will contain not only the deviation and its first derivative, but also second, third, etc. derivatives.

SUMMARY

1. If optimal transients are produced in an ACS with a position-bounded controller the transient duration can be greatly reduced and the dynamic accuracy increased, as compared with ordinary systems.

2. It is difficult to use continuous-action regulators with nonlinear converters to provide optimal transients in practice because the nonlinear links must vary with the type, magnitude and site of the perturbation and with the initial values of the bounded coordinates. Self-oscillation can also occur in such systems.

3. The proposed discrete-action (sampling) regulator can provide optimal transients in response both to step perturbations to load and input, and to uniformly changing input perturbations. The formulas for the controller extreme position dwell time (6) and the equilibrium values of controller coordinates (7) are derived for this regulator.

LITERATURE CITED

[1] Merz, Die Begriffe Schwierigkeit, Leistungsfähigkeit and Durchführbarkeit in der Regelungstechnik. Regelungstechnik, No. 2, 3, 1954.

[2] A. Ia. Lerner, "The limiting speed of an automatic control system," Automation and Remote Control (USSR) 15, 6 (1954).

[3] A. A. Fel'dbaum, "Optimal processes in automatic control systems," Automation and Remote Control (USSR) 14, 6 (1953).

[4] L. Silva, "Predictor servomechanisms", Trans. IRE (March, 1954).

[5] I. G. Ziegler and N. B. Nichols, Optimum Settings for Automatic Controllers, Trans. ASME. vol. 64 (1942).

[6] W. Oppelt, Kleines Handbuch technische Regelvorgänge. GM B. H. Weinheim, 1954.

[7] E. K. Krug and O. M. Minina, "Effect of nonlinear characteristics in an isodrome regulator on the dynamic properties of an automatic control system," Otchet (Report) to the Institute of Automation and Remote Control, Acad. Sci. USSR (1956).

Received March 25, 1957

EVALUATION OF THE ACCURACY OF INPUT REPRODUCTION FOR LINEAR SERVOS AND RECORDING SYSTEMS

V. G. Vasil'ev

(Moscow)

The necessary and sufficient conditions for a linear system to reproduce a given class of inputs accurately are given. The maximum moduli of the inputs and their local rates of change are used to define the class of inputs.

The paper extends earlier ideas by others [1-5].

1. General Concepts and Terminology

1. Reproducing system. By this is meant a device of any operating principle which is intended to transform an input signal into an output, subject to a present relation between their time relationships.

Such systems are subdivided into simple reproducing, differentiating, and integrating, depending on the type of correspondence between input and output.

For a simple system the relation of output $\alpha_{out}(t)$ to input $\alpha_{in}(t)$ is described either by

$$\alpha_{out}(t) = K\alpha_{in}(t), \tag{1}$$

which expresses an idealized followup process, or by

$$\alpha_{out}(t) = K\alpha_{in}(t+\tau), \tag{2}$$

which expresses an idealized recording process.

A system which produces the desired correspondence absolutely precisely will be termed a precise reproducing system. Hence one which effects (1) will be a precise servo system of scale factor K, while one which effects (2) will be a precise recording system of scale factor K and time displacement τ.

Strictly speaking it is impossible to produce a precise system to operate on arbitrary inputs in practice. The reason is that there are various sources of reproduction error.

Here we consider the reproducing properties of linear systems with lumped constant parameters. The reproduction errors in such systems arise from their selective reactions to the inputs.

2. Transfer and distortion functions. The transfer functions for systems of this class belong to the class of rational functions of the complex variable p.

In general, the transfer function of a linear reproducing system K(p), can be represented as the quotient of two polynomials in p:

$$K(p) = K(0)\frac{a(p)}{b(p)} = K(0)\frac{1 + a_1p + \cdots + a_mp^m}{1 + b_1p + \cdots + b_np^n}. \tag{3}$$

To this correspond a precise servo system of transfer function

$$K_0(p) = K(0) \qquad (K(0) = \lim_{p\to 0} K(p)). \tag{4}$$

and a precise recording one of transfer function

$$K_\tau(p) = K(0)e^{p\tau} \quad \left(\tau = \frac{1}{K(0)} \lim_{p\to 0} \frac{d}{dp} K(p) = \frac{K^{(1)}(0)}{K(0)}\right). \tag{5}$$

It is clear that the natures assumed for $K_0(p)$ and $K_\tau(p)$ are such that they can be found exactly by calculation or experiment.

In fact K(0) appearing in (4) and (5) is numerically equal to the output produced by a real system from the input

$$\alpha_{in}(t) = 1(t) \qquad (t = \infty),$$

and $K^{(1)}(0)/K(0)$ is numerically equal to the slope of the phase-frequency characteristic $\varphi(\omega)$ at $\omega = 0$.

It is also advantageous to use some accessory functions (distortion functions) to specify the reproduction properties of the system, namely:

$$\varkappa_0(p) = \frac{K(p) - K_0(p)}{K(0)}, \tag{6}$$

$$\varkappa_\tau(p) = \frac{K(p) - K_\tau(p)}{K(0)}. \tag{7}$$

By using the distortion function we can consider the real system as two independent systems acting in parallel, one being a precise system of transfer function $K_0(p) = K(0)$ (or $K_\tau(p) = K(0)\, e^{\frac{K^{(1)}(0)}{K(0)}p}$) and the other a distortion system of transfer function $\Delta_0(p) = K(0)\varkappa_0(p)$ [or $\Delta_\tau(p) = K(0)\varkappa_\tau(p)$].

3. The input. The law followed by the input is usually unknown. Certain aspects of the class of inputs are frequently known, however.

The first is that all inputs in the class satisfy the Dirichlet conditions in the interval $0 \le t \le \infty$, plus the condition

$$\alpha_{in}(t) \equiv 0 \text{ for } 0 > t \geqslant -\infty.$$

If these conditions are fulfilled it is possible to effect a Laplace transform for each of the inputs:

$$A_{in}(p) = \int_0^\infty \alpha_{in}(t)\, e^{-pt}\, dt.$$

Most frequently, however, the properties of the input are expressed via more concrete and practically convenient formulas.

Thus in the usual frequency-spectrum interpretation* the properties are expressed in terms of the maximum frequency of the spectral components (in that class) and the limiting (for that class) modulus of the inputs:

$$M = \max \sup_{0<t<\infty} |\alpha_{in}(t)|. \tag{8}$$

* This interpretation assumes that one can effect a Fourier transform

$$A_{in}(j\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \alpha_{in}(t)\, e^{-j\omega t}\, dt$$

for any of the probable inputs in the class.

Some [6-8] have used the basic functions of random process theory (correlation and spectral density) to specify the characteristics of the class of inputs.

In the above cases the suitability of the system for reproducing a given class of inputs is analyzed via the frequency characteristics, which illustrate the behavior of the transfer function along the imaginary axis in the p plane.

Other methods are possible. Thus in [2, 3] a complex-spectral interpretation of the input is used;* in addition to M above, the coordinates and dimensions of the region where the poles in the Laplace representation of the probable inputs in the p plane lie, are used.

Here we use M and Ω to express the properties of a given class of inputs.

Parameter Ω is given by

$$\Omega = \max_{0<k<\infty} \sup \sqrt[k]{\sup_{0<t<\infty}\left|\frac{\alpha_{in}^{(k)}(t)}{M}\right|} \tag{9}$$

and is henceforth termed the limiting local rate of change index (for that class); it applies to the time interval $0 \le t \le \infty$.**

Values of Ω are given below for certain elementary types of input.

Type of input	$1(t)$	$e^{\pm\alpha t}$	$\sin \omega t$	$\cos \omega t$
Corresponding value of Ω	0	α	ω	ω

As an accessory means of representing a given class of inputs we use the configuration and coordinates of the region where the poles in the transforms lie. This region is henceforth denoted by $\overline{P}_\alpha$, and the corresponding class by the $\overline{P}_\alpha$ class, specified by M and Ω.

The system suitable for reproducing a $\overline{P}_\alpha$ class can be selected from the data on how the transfer (or distortion) function behaves in the $\overline{P}_\alpha$ region. In some cases these data are more readily derived than those for the frequency characteristics.

2. The Necessary Conditions for Precise Reproduction

1. Reproduction structure. We derive the reproduction structure for $\alpha_{in}(t)$ when the transform $A_{in}(p)$ and $K(p)$ are rational functions of p having no poles at $p = \infty$.

In this case the transform of $\alpha_{in}(t)$ is given by

$$A_{out}(p) = K(p)A_{in}(p),$$

* This interpretation assumes that one can effect a Laplace transform for any of the probable inputs in the class, and also that any such transform is a rational function of p, with all its poles lying in a finite region of the p plane.

** In defining Ω it is assumed that at all discontinuities in $\alpha_{in}(t)$ and in its derivatives $\alpha_{in}^{(k)}(t)$ there are two equally likely values:

$$\begin{aligned} \alpha_{in}^{(k)}(t)_1 &= \lim_{+\Delta t \to 0} \alpha_{in}^{(k)}(t + \Delta t) \\ \alpha_{in}^{(k)}(t)_2 &= \lim_{-\Delta t \to 0} \alpha_{in}^{(k)}(t - \Delta t) \end{aligned} \qquad (k \geqslant 0)$$

and the output $\alpha_{out}(t)$ itself by the Riemann-Mellin formula

$$\alpha_{out}(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} A_{out}(p)\, e^{pt} dp.\ ^{*} \tag{10}$$

Since $A_{out}(p)$ has no special points other than the poles lying in a finite area of the p plane, (10) can be represented as a sum of residues for the poles of $A_{out}(p)e^{pt}$:

$$\alpha_{out}(t) = \sum_{i=z+1}^{N_1} \frac{1}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [A_{in}(p)\, K(p)\, (p+\alpha_i)^{r_i} e^{pt}] +$$
$$+ \sum_{k=z+1} \frac{1}{(s_k-1)!} \lim_{p\to-\beta_k} \frac{d^{s_k-1}}{dp^{s_k-1}} [A_{in}(p)\, K(p)\, (p+\beta_k)^{s_k} e^{pt}] +$$
$$+ \sum_{j=1}^{z} \frac{1}{(q_j-1)!} \lim_{p\to-\gamma_j} \frac{d^{q_j-1}}{dp^{q_j-1}} [A_{in}(p)\, K(p)\, (p+\gamma_j)^{q_j} e^{pt}]. \tag{11}$$

Here α_i is a pole of $A_{in}(p)$ of order r_i, β_k is a pole of K(p) of order s_k, γ_j is a pole common to $A_{in}(p)$ and K(p) of order q_j (for z values i and k: $q_j = r_i + s_k$), N_1 is the number of poles of $A_{in}(p)$, N_2 those of K(p) and z the number of poles common to $A_{in}(p)$ and K(p).

Comparing (11) with how precise servo and recording systems reproduce $\alpha_{in}(t)$, we have

$$\alpha_{0\,out}(t) = \sum_{i=1}^{N_1} \frac{K(0)}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [A_{in}(p)\,(p+\alpha_i)^{r_i} e^{pt}],$$

$$\alpha_{\tau\,out}(t) = \sum_{i=1}^{N_1} \frac{K(0)}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} \left[A_{in}(p)\,(p+\alpha_i)^{r_i} e^{p\left[t+\frac{K^{(1)}(0)}{K(0)}\right]} \right].$$

It is readily shown that the required correspondence between $\alpha_{out}(t)$ and $\alpha_{in}(t)$ can only be provided in a real system via the first term in (11). This term will henceforth be termed the forced (constrained) output component and denoted by $\alpha_{out}(t)_c$.

The second term in (11) describes the free motion of the system, and is hence denoted by $\alpha_{out}(t)_f$.

Finally the third term in (11) expresses the complex resonance state** to which the system goes over on applying $\alpha_{in}(t)$ to the input; it will be termed the resonance component, and denoted by $\alpha_{out}(t)_{res}$.

Hence

$$\alpha_{out}(t) = \alpha_{out}(t)_c + \alpha_{out}(t)_f + \alpha_{out}(t)_{res}. \tag{12}$$

2. Reproduction error. The reproduction properties are evaluated in terms of the deviation between the reproductions of $\alpha_{in}(t)$ by the real system and by a precise reproducing system.

For a servo the deviation is

$$\Delta_0(t) = \alpha_{out}(t) - \alpha_{0\,out}(t)$$

which we term the instantaneous followup error, while for a recording system

$$\Delta_\tau(t) = \alpha_{out}(t) - \alpha_{\tau\,out}(t)$$

* In this case σ must be greater than the real part of any of the poles of $A_{out}(p)$.

** Strelkov [2, 3] first used the term 'complex resonance'. The ordinary frequency resonance is a particular case.

is the instantaneous recording error.

Hence from (12)

$$\Delta_0(t) = [\alpha_{out}(t)_c - \alpha_{0\,out}(t)] + \alpha_{out}(t)_f + \alpha_{out}(t)_{res}, \tag{13a}$$

$$\Delta_\tau(t) = [\alpha_{out}(t)_c - \alpha_{\tau\,out}(t)] + \alpha_{out}(t)_f + \alpha_{out}(t)_{res}. \tag{13b}$$

The first term in (13a) is the forced component of the followup error $\Delta_0(t)_c$, the first in (13b) the forced component of the recording error $\Delta_\tau(t)_c$.

(13a) and (13b) imply that the free and resonance components of the followup and recording errors are identical, i.e.

$$\Delta(t)_f = \Delta_0(t)_f = \Delta_\tau(t)_f = \alpha_{out}(t)_f,$$
$$\Delta(t)_{res} = \Delta_0(t)_{res} = \Delta_\tau(t)_{res} = \alpha_{out}(t)_{res}.$$

It is readily shown from the structure of $\Delta_0(t)$ and $\Delta_\tau(t)$ that the components $\Delta_{0(\tau)}(t)_c$, $\Delta_{0(\tau)}(t)_f$ and $\Delta_{0(\tau)}(t)_{res}$ differ markedly from one another in orgin and in time variation.

It is thus desirable to examine each error component (followup or recording) separately.

As metrological criteria for evaluating the reproduction accuracy for a given class of inputs we can use the maximum values of the Tchebycheff criteria in the class:

$$\widetilde{\Delta}_{0(\tau)} = \sup_{0<t<\infty} |\Delta_{0(\tau)}(t)_c|, \text{ i. e. } \Delta_{0(\tau)} = \max \widetilde{\Delta}_{0(\tau)c}, \tag{14a}$$

$$\widetilde{\Delta}_f = \sup_{0<t<\infty} |\Delta_{0(\tau)}(t)_c|, \text{ i. e. } \Delta_f = \max \widetilde{\Delta}_f, \tag{14b}$$

$$\widetilde{\Delta}_{res} = \sup_{0<t<\infty} |\Delta_{0(\tau)}(t)_c|, \text{ i. e. } \Delta_{res} = \max \widetilde{\Delta}_{res}, \tag{14c}$$

and also the minimum damping rate $\Delta(t)_f = \alpha_{out}(t)_f$.*

3. <u>Subdivision of the conditions for precise reproduction.</u> The nature of the correspondence between the class of inputs and the system suitable for reproducing them determine the conditions for precise reproduction.

We must here differentiate the necessary from the sufficient conditions for precise reproduction.

The necessary establish that the system is in principle suitable for the given class of inputs.

The sufficient establish that the system is suitable for reproducing the given class of inputs with a set accuracy.

Two types of condition (precise followup and precise recording) can be differentiated in terms of the type of correspondence required between input and output.

---

* From the practical viewpoint two factors can with advantage be used to specify the free motion:

1. The Tchebycheff criterion $M_f = \sup_{0 \le t \le \infty} |\bar{\alpha}_{out}(t)_f|$, being the maximum value of the modulus to the input $\alpha_{in}(t) = M1(t)$ in the interval $0 \le t \le \infty$.

2. The response time of the free motion, T_f, i.e. the time during which $|\bar{\alpha}_{out}(t)_f|$ does not exceed some preset value $K(0)M\bar{\epsilon}_f$:

$$K(0)M\bar{\epsilon}_f = \alpha_{out}(T_f)_f = \sup_{T_f < t < \infty} |\alpha_{out}(t)_f| \le M_f.$$

4. The first of the necessary conditions for precise reproduction. (11)can be used to show that simple input reproduction is possible only in a stable reproducing system.* We can only speak of more or less precise reproduction for a given class of inputs when the rate at which the free motion dies away substantially exceeds the rate at which any of the inputs does. This is the first of the necessary conditions for precise reproduction (follow-up or recording) for a given class of inputs. If the line Re(p) = −a− is the right boundary of the region of special points of K(p) (region $\overline{P}_\beta$ in the figure) and Re(p) = − b − is the left boundary to the region $\overline{P}_\alpha$, the first of the necessary conditions for precise reproduction can be put as

$$-a \ll -b. \tag{15}$$

If this is fulfilled in region $\overline{P}_\alpha$ there are no poles in the transfer system (z = 0), so no complex resonance occurs with any of the probable inputs in the class. Thus when the first of the necessary conditions for precise reproduction is fulfilled

$$\Delta_{res} = \max \sup_{0<t<\infty} |\Delta(t)_{res}| = 0,$$

$$\Delta_{0(\tau)}(t) = \Delta_{0(\tau)}(t)_C + \Delta_{0(\tau)}(t)_f\,.$$

5. The second necessary condition. Mere fulfillment of the first necessary condition does not ensure that the system is in principle entirely suitable for reproducing the given class of inputs. This only occurs if, together with (15) being fulfilled, the Tchebycheff criteria

$$\widetilde{\Delta}_{0\,C} = \sup_{0<t<\infty} |\Delta_0(t)_C|, \quad \widetilde{\Delta}_{\tau\,C} = \sup_{0<t<\infty} |\Delta_\tau(t)_C|$$

are suffiently small for the given class of inputs.

Since when z = 0

$$\Delta_0(t)_C = \sum_{i=1}^{N_1} \frac{K(0)}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [\varkappa_0(p) A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}], \tag{a}$$

$$\Delta_\tau(t)_C = \sum_{i=0}^{N_1} \frac{K(0)}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [\varkappa_\tau(p) A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}], \tag{b}$$

it is readily seen that $\Delta_0(t)_C$ and $\Delta_T(t)_C$ will be sufficiently small only if the moduli of $\kappa_0(p)$ and $\kappa_T(p)$ and of their first $(r_i - 1)$ derivatives in region $\overline{P}_\alpha$ are sufficiently small.

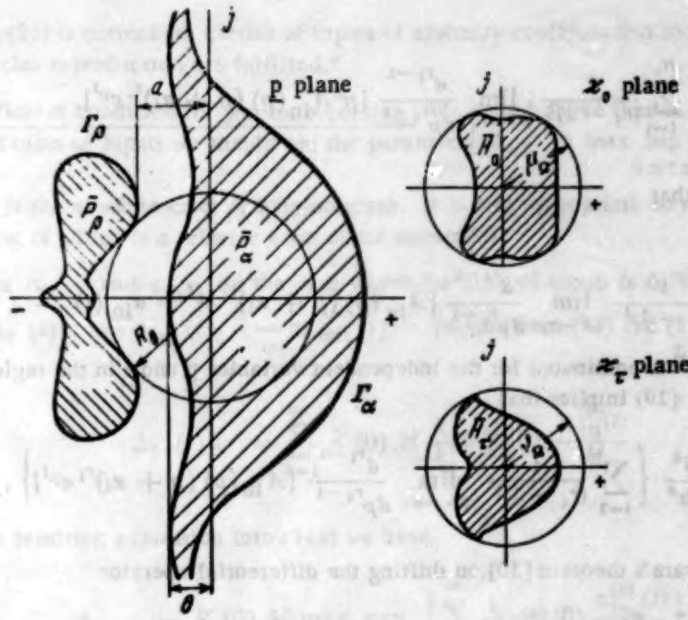
This is the second necessary condition for precise reproduction.

Let us delineate region $\overline{P}_\alpha$ in some regions $\overline{H}_0$ and $\overline{H}_T$ in the distortion planes $\kappa_0 = \kappa_0(p)$ and $\kappa_T = \kappa_T(p)$, respectively (see Figure).

Formulas (a) and (b) above show that the minimum $\Delta_{0\,C}$ and $\Delta_{T\,C}$ correspond to $\kappa_0(p)$ and $\kappa_T(p)$ being uniformly continuous in some closed region containing region $\overline{P}_\alpha$, and to there being no poles in the transfer function in the region $\overline{P}_\alpha$.

This in turn means that a necessary condition is that the transform of the $\overline{P}_\alpha$ region onto the appropriate distortion planes should be compact. The less the radius μ_Ω of the circle enclosing $\overline{H}_T$ the less $\Delta_{0\,C}$ for the given class of inputs; the less the radius ν_Ω of the circle enclsoing $\overline{H}_T$ the less $\Delta_{T\,C}$.

* Korn first discovered this as a required element in a problem of his on integral synchronization, i.e., a problem of precise reproduction. Kharkevich [5] has given a fairly comprehensive review of the papers on distortion in reproduction.



3. The First Sufficient Condition for Precise Reproduction - Precise Followup

1. Structure of the forced component in the output. We assume the class of inputs and the reproducing system to comply with the necessary conditions. Then z = 0; hence

$$\alpha_{\text{out}}(t)_c = \sum_{i=1}^{N_1} \frac{1}{(r_i-1)!} \lim_{p \to -\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [K(p) A_{\text{in}}(p) (p+\alpha_i)^{r_i} e^{pt}]. \tag{16}$$

We also suppose that the $\bar{P}_\alpha$ region lies in the uniform convergence zone of the series

$$\sum_{k=0}^{\infty} \frac{1}{k!} K^{(k)}(0) p^k, \tag{17}$$

which represents the transfer function for the system within the circle

$$|p| < R = \frac{1}{\overline{\lim\limits_{k\to\infty}} \sqrt[k]{|K^{(k)}(0)|}}.$$

A uniformly converging series can be differentiated an arbitrary number of times. Hence we can substitute (17) into (16):

$$\alpha_{\text{out}}(t)_c = \sum_{i=1}^{N_1} \frac{1}{(r_i-1)!} \times$$

$$\times \lim_{p \to -\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} \left[\sum_{k=0}^{\infty} \frac{1}{k!} K^{(k)}(0) A_{\text{in}}(p) (p+\alpha_i)^{r_i} e^{pt} p^k \right].$$

Supposing it permissible to change the order of summation, we have

$$\alpha_{\text{out}}(t)_c = \sum_{k=0}^{\infty} \frac{1}{k!} K^{(k)}(0) \sum_{i=1} \frac{1}{(r_i-1)!} \times \tag{18}$$

$$\times \lim_{p \to -\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [p^k A_{\text{in}}(p) (p+\alpha_i)^{r_i} e^{pt}].$$

The meaning of

$$\sum_{i=1}^{N_1} \frac{1}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [p^k A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}],$$

is revealed if we recollect that

$$\sum_{i=1}^{N_1} \frac{1}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}] = \alpha_{in}(t) \qquad (19)$$

and that $A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}$ is continuous for the independent variables p and t in the region of some point α_i for the interval $0 \leq t \leq \infty$. (19) implies that

$$\alpha^{(k)}_{in}(t) = \frac{d^k}{dt^k} \left\{ \sum_{i=1} \frac{1}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}] \right\},$$

or, in accordance with Schwarz's theorem [10], on shifting the differential operator

$$\alpha^{(k)}_{in}(t) = \sum_{i=1}^{N_1} \frac{1}{(r_i-1)!} \lim_{p\to-\alpha_i} \frac{d^{r_i-1}}{dp^{r_i-1}} [p^k A_{in}(p)(p+\alpha_i)^{r_i} e^{pt}].$$

The resulting expression enables us to put (18) into the form used in servo systems theory [6, 7]:

$$\alpha_{in}(t)_c = \sum_{k=0}^{\infty} \frac{1}{k!} K^{(k)}(0)\, \alpha^{(k)}_{in}(t). \qquad (20)$$

We shall prove that this formula is correct for a class of inputs more extensive than that used in its derivation.

Suppose that we apply to the input of a system with the transfer function of (3) an arbitrary bounded input $\beta_{in}(t)$ which has all its derivatives finite for $0 \leq t \leq \infty$. If the initial conditions are zero the forced component in the output $\beta_{out}(t)_c$ is a particular solution to

$$\beta_{out}(t)_c + b_1\beta^{(1)}_{out}(t)_c + \ldots + b_n\beta^{(k)}_{out}(t)_c =$$
$$= K(0)\,\beta_{in}(t) + \ldots + K(0)\,a_m\beta^{(m)}_{in}(t).$$

This solution is sought in the form of a series

$$\beta_{out}(t) = \sum_{k=0}^{\infty} C_k \beta^{(k)}_{in}(t),$$

which can be differentiated n times.

Substituting this series into the differential equation shows that when

$$C_k = \frac{1}{k!} K^{(k)}(0)$$

it becomes an identity.*

---

* To prove this it is best to use (24a) and (24b) below.

This proves that (20) is correct for classes of inputs of arbitrary configuration in the $\overline{P}_\alpha$ region if the necessary conditions for precise reproduction are fulfilled.*

2. <u>The first sufficient condition for precise recording.</u> We stated above that in evaluating the forced error component for a given class of inputs we should use the parameter $\Delta_{0\,C} = \max \sup\limits_{0 \le t \le \infty} |\Delta_0(t)_C|$, the metrological analog of which is the absolute error of measurement. It is also appropriate to use parameter $\epsilon_{0C} = \dfrac{\Delta_{0C}}{K(0)M}$, the metrological analog of which is a relative error of measurement.

We now evaluate Δ_{0C} and ϵ_{0C} for the case where the class of inputs is defined by Ω and M.

By definition $\Delta_0\,(t)_C = \alpha_{out}(t)_C - \alpha_{0\,out}(t)$; hence, from (20) and (1), we can put

$$\Delta_0\,(t)_C = \sum_{k=1}^{\infty} K(0)\,M \frac{1}{k!} \varkappa_0^{(k)}(0) \frac{\alpha_{in}^{(k)}(t)}{M}.$$

Substituting the resulting expression into (14a) we have

$$\Delta_{0\,C} = K(0)\,M \max \sup_{0<t<\infty} \left| \sum_{k=1}^{\infty} \frac{1}{k!} \varkappa_0^{(k)}(0) \frac{\alpha_{in}^{(k)}(t)}{M} \right|.$$

Evaluating Δ_{0C}, we get

$$\Delta_{0\,C} \leqslant K(0)\,M \sum_{k=1}^{\infty} \left| \frac{1}{k!} \varkappa_0^{(k)}(0) \right| \max \sup_{0<t<\infty} \left| \frac{\alpha_{in}^{(k)}(t)}{M} \right|. \tag{21}$$

Comparing $\max \sup\limits_{0 \le t \le \infty} \left| \dfrac{\alpha_{in}^{(k)}(t)}{M} \right|$ with (9) we see that

$$\max \sup_{0<t<\infty} \left| \frac{\alpha_{in}^{(k)}(t)}{M} \right| \leqslant \Omega^k.$$

This enables us to put (21) as

$$\Delta_{0C} \leqslant K(0)\,M \sum_{k=1}^{\infty} \frac{1}{k!} \,|\, \varkappa_0^{(k)}(0) \,|\, \Omega^k,$$

from which**

$$\varepsilon_{0\,C} \leqslant \sum_{k=1}^{\infty} \frac{1}{k!} \,|\, \varkappa_0^{(k)}(0) \,|\, \Omega^k. \tag{22}$$

* It can also be proved that (20) is correct for linear systems with distributed constants.

** It can be shown that, when not less than the first <u>m</u> derivatives in the given class of inputs are defined, we can use to estimate ϵ_{0C} an inequality

$$\varepsilon_{0\,C} - |\,\varkappa_0^{(1)}(0)\,|\,\lambda_1 + \ldots + \frac{1}{(m-1)!}\,|\,\varkappa_{(0)}^{(m-1)}(0)\,|\,\lambda_{m-1} + \frac{1}{m!}\sigma_m\lambda_m. \tag{22a}$$

Here $\lambda_i M$ is the limiting value (for the class) of the i-th derivative of the input in the range $0 < t \le \infty$;

$$\sigma_m = \frac{1}{2\pi}\int_0^\infty \tau^m \left| \int_{-j\infty}^{j\infty} K(p)\, e^{pt} dp \right| d\tau.$$

Equation (22) is clearly a special case of (22a), since m = ∞ and $\lim\limits_{m\to\infty} \sigma_m = \lim\limits_{m\to\infty} |\,\varkappa_0^{(m)}(0)\,|$.

Certain difficulties are encountered in using (22); we have to compute the series

$$\sum_{k=0}^{\infty} \frac{1}{k!} \mid \varkappa_0 (0) \mid \Omega^k. \tag{23}$$

Hence, practically speaking, in determining ϵ_{0C} it is best to use inequalities such as $\epsilon_{0C} \leq \eta_i$, where η_i is one of the calculable majorants of (23).

The majorants $\{\eta_i\}$ enable one to express the first sufficient condition for precise reproduction in the form $\eta_i \leq \epsilon\bar{\epsilon}_C$, where $\epsilon\bar{\epsilon}_C$ is some preset value.

Compliance with this inequality is sufficient to ensure that the relative deviation of $\alpha_{out}(t)_C$ from $\alpha_{out}(t)$ for any of the inputs in the given class will not during time $0 \leq t \leq \infty$ exceed $\epsilon_{\overline{0C}}$.

Strictly speaking, we can construct an infinite number of calculable majorants to (23), and hence an infinite number of forms for the first sufficient condition. But in practice only those majorants are of interest which have sufficiently simple formulas when (23) is evaluated sufficiently accurately.

3. The majorants of $\sum_{k=1}^{\infty} \frac{1}{k!} \mid \varkappa_0^{(k)} (0) \mid \Omega^k$. We will derive the majorants for certain cases of determining $\kappa_0(p)$ and $K(p)$.

Let the distortion function be

$$\varkappa_0 (p) = \frac{(a_1 - b_1)\, p + (a_2 - b_2)\, p^2 + \ldots}{1 + b_1 p + \ldots + b_n p^n}$$

defined by the values

$$|\, b_1 \,|, \ldots, |\, b_n \,|, \ |\, a_1 - b_1 \,|, \ |\, a_2 - b_2 \,|, \ldots$$

It can be shown [12] that the coefficients in the Maclaurin series of the distortion function $B_k = \frac{1}{k!} \varkappa_0^{(k)} (0)$ are related by recurrence formulas

$$B_k = \Delta_1 A_{k-1} + \Delta_2 A_{k-2} + \ldots, \tag{24a}$$

$$A_k = - b_1 A_{k-1} - b_2 A_{k-2} + \ldots, \tag{24b}$$

where

$$\Delta_i = a_i - b_i, \qquad A_i = \begin{cases} 0 & \text{for} \quad i < 0, \\ 1 & \text{for} \quad i = 0, \end{cases}$$

$$a_i = 0 \quad \text{for} \quad i > m, \quad b_i = 0 \quad \text{for} \quad i > n.$$

(24a) enables us to represent (23) as an infinite sum $\sum_{k=1}^{\infty} \mid B_k \mid \Omega^k$, each term in which can be evaluated from

$$\begin{aligned} &|\, B_1 \,|\, \Omega = |\, \Delta_1 \,|\, \Omega, \\ &|\, B_2 \,|\, \Omega^2 \leqslant |\, \Delta_1 \,|\,|\, A_1 \,|\, \Omega^2 + |\, \Delta_2 \,|\, \Omega^2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &|\, B_k \,|\, \Omega^k \leqslant |\, \Delta_1 \,|\,|\, A_{k-1} \,|\, \Omega^k + |\, \Delta_2 \,|\,|\, A_{k-2} \,|\, \Omega^k + \ldots + |\, \Delta_k \,|\, \Omega^k. \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

We thus get

$$\epsilon_{0\,C} \leqslant \sum_{k=1}^{\infty} |\, B_k \,|\, \Omega^k \leqslant \sum_{i=1}^{h} |\Delta_i|\, \Omega^i \left[1 + \sum_{k=1}^{\infty} |\, A_k \,|\, \Omega^k \right],$$

$$(n < l \geqslant m, \quad a_i = 0 \quad \text{for} \quad i > m, \quad b_i = 0 \ \text{for} \quad i > n). \tag{25}$$

Remembering that when $a_1 = a_2 = \ldots = a_m = 0$, $B_k = A_k$ we can, by using (24b), put

$$\sum_{k=1}^{\infty} |A_k| \Omega^k \leqslant \sum_{i=1}^{k} |\Delta_i| \Omega^i \left[1 + \sum_{k=1}^{\infty} |A_k| \Omega^k\right],$$

from which

$$\sum_{k=1}^{\infty} |A_k| \Omega^k \leqslant \frac{\sum_{i=1}^{n} |b_i| \Omega^i}{1 - \sum_{i=1}^{n} |b_i| \Omega^i}.$$

Substituting the estimate of $\sum_{k=1}^{\infty} |A_k| \Omega^k$ into (25) we get

$$\epsilon_{0\,C} \leqslant \frac{\sum_{i=1}^{l} |\Delta_i| \Omega^i}{1 - \sum_{i=1}^{n} |b_i| \Omega^i} = \eta_1 \tag{26}$$

Then the ϵ_{0C} is given the best for the class of inputs defined by Ω and M, and for the set of reproducing systems defined by the values $|b_1|, \ldots, |b_n|, |a_1 - b_1|, |a_2 - a_2|$ ...,etc.

Similarly, it is easily shown that

$$\epsilon_{0\,C} = \frac{\sum_{i=1}^{l} |\Delta_i| \Omega^i}{1 - \sum_{i=1}^{n} |b_i| \Omega^i} = \eta_1 ,$$

if

$$\alpha_{\text{in}}(t) = Me^{-\Omega t},$$

$$\varkappa_0(p) = \frac{\sum_{i=1}^{l} (-1)^i |\Delta_i| p^i}{1 - \sum_{i=1}^{n} (-1)^i |b_i| p^i}.$$

Analysis of $\eta_1(\Omega)$ shows that it gives (as well as $\epsilon_{0\,C}$) an estimate of the maximum values of $|\kappa_0(p)|$ in the circle $|p| \leq \Omega$: for the given class of inputs:

$$|\varkappa_0(p)| \leqslant \mu_\Omega \leqslant \frac{\sum_{i=1}^{l} |\Delta_i| \Omega^i}{1 - \sum_{i=1}^{n} |b_i| \Omega^i}.$$

This implies that

$$\epsilon_{0C} \approx \mu_\Omega^{*}. \tag{27}$$

* There is every reason to suppose that $\epsilon_{0C} \approx \mu_\Omega$ retains its validity when $\kappa_0(p)$ belongs to a class of functions more extensive than the rational ones used here.

In particular, we can prove it correct when $\kappa_0(p)$ is a holomorphic function of a single sheet within the circle $|p| \leqslant R_b = \dfrac{1}{\overline{\lim\limits_{k\to\infty}} \sqrt[k]{|B_k|}}$ which contains the $\overline{P}_\alpha$ region.

Estimates of (23) of forms other than that of (26) can be made using combined parameters for the reproducing system.

Such parameters are*

$$R_b = \frac{1}{\overline{\lim\limits_{1<k\to\infty}} \sqrt[k]{\frac{1}{k!}\left|x_0^{(k)}(0)\right|}},$$

$$R_\Delta = \frac{1}{\lim\limits_{1\leq k\to\infty} \sqrt[k]{\frac{1}{k!}\left|\left(\frac{b(p)}{1+a(p)-b(p)}\right)^k_{p=0}\right|}},$$

$$D_{0(\tau)} = \sup_{1<k<\infty} \sqrt[k]{\frac{1}{k!}\left|x^{(k)}_{0(\tau)}(0)\right|},$$

$$Q_{0(\tau)} = \sup_{1<k<\infty} \sqrt{\left|x^{(k)}_{0(\tau)}(0)\right|},$$

$$L = \frac{1}{K(0)} \sup_{1\leq k\leq N} \left\{ \sup_{1\leq i\leq s_k} \frac{1}{(s_k-i)!} \lim_{p\to-\beta_k} \frac{d^{s_k-i}}{dp^{s_k-i}} \left[K(p)(p+\beta_k)^{s_k}\right]\right\}.$$

We now derive the majorants of interest to us by using the grouped parameters enumerated above.

a) Let the distortion function be defined by R_b and R_Δ, and by the numbers m and n.

By definition, R_b is the distance between $p = 0$ and the nearest zero of $b(p)$, while R_Δ is the distance between $p = 0$ and the nearest zero of $[\Delta(p) + 1] = \delta(p)$.

It is readily shown that the coefficients in $b(p)$ and $\delta(p)$ are related to R_b and R_Δ by inequalities of the type

$$|b_i| \leq C_n^i \frac{1}{R_b^i}, \quad |\Delta_i| \leq C_l^i \frac{1}{R_\Delta^i}.$$

Consequently,

$$\sum_{i=1}^{n} |b_i|\,\Omega^i \leq \sum_{i=1}^{n} C_n^i \frac{\Omega^i}{R_b^i} = \left(1+\frac{\Omega}{R_b}\right)^n - 1,$$

$$\sum_{i=1}^{l} |\Delta_i|\,\Omega^i \leq \sum_{i=1}^{l} C_l^i \frac{\Omega^i}{R_\Delta^i} = \left(1+\frac{\Omega}{R_\Delta}\right)^n - 1.$$

Substituting these in (26) we get

$$\varepsilon_{0\,c} \leq \frac{\left(1+\frac{\Omega}{R_\Delta}\right)^l - 1}{2-\left(1+\frac{\Omega}{R_b}\right)^n} = \eta_2 \quad . \tag{28}$$

* The parameters R_b and R_Δ, to determine which Cauchy-Adamar fomula is used [9], are the convergence radii of the Maclaurin series for $\kappa_0(p) = \frac{\Delta(p)}{b(p)}$ and $\frac{b(p)}{\Delta(p)+1} = \frac{b(p)}{\delta(p)}$, respectively. It can be shown that if the coefficients in the polynomials $b(p)$ and $\delta(p)$ are everywhere positive, it is permissible to replace R_b and R_Δ in all the estimates of (23) given below by the lower bounds to the zeros of $b(p)$ and $\delta(p)$, respectively. From a certain theorem [11] the bounds are

$$\lambda_b = \left[\sup_{0<i\leq n} \frac{b_i}{b_{i-1}}\right]^{-1} \text{ and } \lambda_\Delta = \left[\sup_{0<i\leq l} \frac{\Delta_i}{\Delta_{i-1}}\right]^{-1}$$

b) Assume a distortion function defined by D_0 (a parameter).

By definition, $|B_k| \leq D_0^k$; hence

$$\varepsilon_{0\,c} \leqslant \sum_{k=1}^{\infty} D_0^k \Omega^k = \frac{\Omega D_0}{1-\Omega D_0} = \eta_3 \quad . \tag{29}$$

If the distortion function is defined by the coefficients in a(p) and b(p) the B_k needed to determine D_0 can be found from

$$B_k = (-1)^{k-1} \begin{vmatrix} \Delta_1 & \Delta_2 & \Delta_3 & \cdots & \Delta_k \\ 1 & b_1 & \cdots & & \cdot \\ 0 & \cdot & \cdot & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & \cdot & b_3 \\ & & \cdots & b_1 & b_2 \\ 0 & \cdots & 0 & 1 & b_1 \end{vmatrix} = (-1)^{k-1} D_k, \; (1 \leqslant k \neq 0),$$

where $\Delta_i = 0$ when $i > l$, $b_i = 0$ when $i > n$, $a_i = 0$ when $i > m$.

c) Letting the distortion function be defined by Q_0 we can write

$$|x_0^{(k)}(0)| \leqslant Q_0^k,$$

$$\varepsilon_{0\,c} \leqslant \sum_{k=1}^{\infty} \frac{1}{k!} Q_0^k \Omega^k = e^{Q_0 \Omega} - 1 = \eta_4 \tag{30}$$

Q_0 may also be found from

$$Q_0 = \sup_{1<k<\infty} \sqrt[k]{k!\,|D_k|}.$$

d) Let the transfer function be given via the parameters L and k_b, and also by the numbers of poles by orders. By definition

$$K(p) = LK(0)\left[\frac{x_1}{p+\beta_1} + \cdots + \frac{x_{s_1}}{(p+\beta_1)^{s_1}} + \cdots \right.$$

$$\left. \cdots + \frac{x_{s_1+\cdots+s_{N-1}+1}}{(p+\beta_N)} + \cdots + \frac{x_{s_1+\cdots+s_N}}{(p+\beta_N)^{s_N}}\right],$$

where $\sup\limits_{1 \leq i \leq s_1 + \cdots + s_N = n} |\varkappa|_i = 1$, N is the total number of poles of K(p) and s_i is the order for pole β_i.

Hence, here, for $k \geq 1$, we have:

$$B_k = \frac{1}{k!} x_0^{(k)}(0) = (-1)^k L \left[C_k^k \left(\frac{x_1}{\beta_1^{k+1}} + \cdots + \frac{x_{s_1+\cdots+s_{N-1}+1}}{\beta_N^{k+1}} \right) + \right.$$

$$+ C_{k+1}^k \left(\frac{x_2}{\beta_1^{k+2}} + \cdots + \frac{x_{s_1+\cdots+s_{N-1}+2}}{\beta_N^{k+2}} \right) +$$

$$\left. + C_{k+2}^k \left(\frac{x_3}{\beta_1^{k+3}} + \cdots + \frac{x_{s_1+\cdots+s_{N-1}+3}}{\beta_N^{k+3}} \right) + \cdots \right].$$

Using the fact that

$$x_{s_1+\dots+s_{j-1}+i}=0 \text{ for } i>s_j, \quad \sup_{1<i<N} s_i = s, \quad \sup_{1<i<N}\left|\frac{1}{\beta_i}\right|=\frac{1}{R_b}$$

$$\left|\left(\frac{x_i}{\beta_1^{k+i}}+\dots+\frac{x_{s_1+s_2+\dots+s_{N-1}+i}}{\beta_N^{k+i}}\right)\right|<\frac{N_i}{R_b^{k+i}}$$

(where N_i is the number of poles of order not less than i) we can put

$$|B_k|<L\left\{C_k^k\frac{N_1}{R_b^{k+1}}+C_{k+1}^k\frac{N_2}{R_b^{k+2}}+\dots+C_{k+s-1}^k\frac{N_s}{R_b^{k+s}}\right\}.$$

Substituting this estimate of $|B_k|$ into (23) we get

$$\sum_{k=1}^{\infty}|B_k|\Omega^k<L\frac{N_1}{R_b}\left[\frac{R_b}{R_b-\Omega}-1\right]+\dots+L\frac{N_s}{R_b^s}\left[\frac{R_b^s}{(R_b-\Omega)^s}-1\right]\geq\epsilon_{0c}$$

This estimate of ϵ_{0c} can be somewhat simplified. In fact $N_1 = N$, $\sup_{1\leq i\leq s} N_i/N_1 \leq 1$; hence

$$\epsilon_{0c}<\frac{LN}{(R_b-\Omega)^s}\frac{(R_b-\Omega)^s-1}{(R_b-\Omega)-1}\left[1-\frac{(R_b-\Omega)^s}{R_b^s}\frac{R_b^s-1}{(R_b-\Omega)^s-1}\frac{(R_b-\Omega)-1}{R_b-1}\right]=\eta_6 \qquad (31)$$

In the special case $s = 1$ we have

$$\epsilon_{0c}\leq\frac{\Omega}{R_b}\frac{LN}{R_b-\Omega}.$$

Analysis of the estimates of ϵ_{0c} shows that they all have certain common properties.

1. All are absolutely precise for the input

$$\alpha_{in}^{-}(t)=M\,1(t) \quad (\Omega=0).$$

2. All are absolutely precise for a precise servo.*

3. Like (26), (28)-(31) are the best for a given class of inputs and for the set of systems defined by the grouped parameters used in the estimates.**

These properties indicate that our estimates of ϵ_{0c} above are compratively of high accuracy. The accuracy is the greater the smaller Ω, the smaller the size of the $\overline{F}_\alpha$ region and the more monotonic the distortion function in this region.

A further property of the estimates of ϵ_{0c} is to be noted.

Considering the majorants of (23) for values of Ω which ensure that $\eta_i(\Omega)\leq\overline{\epsilon}_{0c}$ is complied with, and remembering that in most practical instances $0<\overline{\epsilon}_{0c}\ll 1$ and $|B_1|\Omega \gg \frac{1}{k!}|B_k|\Omega^k$ $(1<k\leq\infty)$, it is readily seen that for any of our marjorants $\eta_i(\Omega)\approx\Omega\lim_{\Omega\to 0}\frac{d}{d\Omega}[\eta_i(\Omega)]$.

* For a precise servo $a_1=a_2=\dots=a_m=0$, $b_1=b_2=\dots=b_n=0$ or $a_i=b_i$ $(i=1,2,\dots,n)$, $R_b=R_\Delta=\infty$ and $D_0=Q_0=L=0$.

** Thus when reproducing $\alpha_{in}(t)=Me^{-\Omega t}$ (a class of inputs specified by M and Ω) with a system of distortion function $\kappa_0(p)=-\frac{pD_0}{1+pD_0}$ (a set of reproducing systems specified by D_0) we have $\sup_{0\leq t\leq\infty}|(t)_c|=\frac{\Omega D_0}{1-\Omega D_0}=\eta_3=\epsilon_{0c}$.

Thus to perform a preliminary check on the correspondence between a given class of inputs and the reproducing system we may use the relationships

$$\varepsilon_{0\,c} \approx |\Delta_1|\,\Omega \leqslant \bar{\varepsilon}_{0\,c}\,, \quad \varepsilon_{0_c} \approx \frac{l}{R_\Delta}\,\Omega \leqslant \bar{\varepsilon}_{0c}$$

$$\varepsilon_{0\,c} \approx D_0\Omega \leqslant \bar{\varepsilon}_{0c}\,, \quad \varepsilon_{0c} \approx Q_0\Omega \leqslant \bar{\varepsilon}_{0\,c}\,, \tag{32}$$

$$\varepsilon_{0\,c} \approx \frac{L}{R_b} N\Omega \leqslant \bar{\varepsilon}_{0c}\,.$$

Of these, one of special interest is *

$$\varepsilon_{0\,c} \approx |\Delta_1|\,\Omega = \frac{1}{1!}\,|\varkappa_0^{(1)}(0)|\,\Omega = |B_1|\,\Omega = \lim_{\omega\to 0}\left|\frac{d}{d\omega}\varphi(\omega)\right|\Omega,$$

from which ϵ_{0c} can be estimated from the experimental phase-frequency response.

4. The First Sufficient Condition for Precise Reproduction-Precise Recording.

1. Structure of $\Delta_\tau(t)_c$ We have already shown that the forced components of the instantaneous recording error can be calculated from

$$\Delta_\tau(t)_c = \alpha_{\text{out}}(t)_c - \alpha_{\tau\,\text{out}}(t). \tag{33}$$

Provided the necessary conditions for precise reproduction are complied with, the first term $\alpha_{\text{out}}(t)_c$ may be shown as the functional series

$$\sum_{k=0}^{\infty} \frac{1}{k!} K^{(k)}(0)\,\alpha_{\text{in}}^{(k)}(t) = \alpha_{\text{out}}(t)_c\ ; \tag{34}$$

while the second, $\alpha_{\tau\,\text{out}}(t) = K(0)\,\alpha_{\text{in}}\left(t + \frac{K^{(1)}(0)}{K(0)}\right)$ may be shown as the functional series

$$K(0)\sum_{k=0}^{\infty} \frac{1}{k!}\left[\frac{K^{(1)}(0)}{K(0)}\right]^k \alpha_{\text{in}}^{(k)}(t) = \alpha_{\tau\,\text{out}}(t). \tag{35}$$

Substituting (34) and (35) into (33) we get

$$\Delta_\tau(t)_c = \sum_{k=0}^{\infty} \frac{1}{k!}\varkappa_\tau^{(k)}(0)\,\alpha_{\text{in}}^{(k)}(t) = \sum_{k=2}^{\infty} C_k \alpha_{\text{in}}^{(k)}(t),$$

where

$$\varkappa_0^{(k)}(0) = \begin{cases} 0 & \text{for } k \leqslant 1 \\ \dfrac{K^{(k)}(0)}{K(0)} - \left[\dfrac{K^{(1)k}(0)}{K(0)}\right]^k & \text{for } k \geqslant 1. \end{cases}$$

---

* There is every reason to assume that $\kappa_0(p)$ retains its validity when k belongs to a class of functions more extensive than the rational ones used here.

In particular, we can prove it correct when $\kappa_0(p)$ is a holomorphic function of a single sheet within the circle $|p| \leqslant R_b \left(R_b = \dfrac{1}{\overline{\lim\limits_{1<k\to\infty}} \sqrt[k]{|B_k|}}\right)$, which contains the $\bar{P}_\alpha$ regions

$$\varepsilon_{0c} \leqslant \Omega\,|B_1|\,\frac{R_b^2}{(R_b - \Omega)^2} \approx \Omega\,|B_1| \quad \text{for} \quad \Omega \ll 1, \quad R_b \gg 1.$$

The proof of this involves using the properties of single-sheet functions which are holomorphic and normalized within the circle $\left|\ z = \dfrac{p}{R_b}\ \right| < 1$.

2. The first sufficient conditon for precise recording. By using a proof similar to that given for (22) we can deduce the following estimate for the limiting value of $\Delta_{\tau\,c}$ for the given class of inputs:

$$\Delta_{\tau\,c} \leqslant K(0)M \sum_{k=0}^{\infty} \frac{1}{k!} |\varkappa_1^{(k)}(0)|\Omega^k \leqslant K(0)M \sum_{k=2}^{\infty} |C_k|\Omega^k.$$

or

$$\varepsilon_{\tau\,c} = \frac{\Delta_{\tau\,c}}{K(0)M} \leqslant \sum_{k=2}^{\infty} |C_k|\Omega^k. \qquad (36)$$

The practical use of (36), just as with (22), involves certain difficulties in calculating the series

$$\sum_{k=2}^{\infty} |C_k|\Omega^k. \qquad (37)$$

These difficulties are removed by replacing series (37) appearing in (36) by some one of its calculable majorants. Let ξ_1 be one of these. Then the first sufficient condition for precise recording can be put as $\epsilon_{\theta c} \leq \leq \xi_1 \leq \bar{\varepsilon}_{\tau\,c}$, where $\bar{\varepsilon}_{\tau\,c}$ is some preset number.

Compliance with this is sufficient to ensure that the system will reproduce any of the probable inputs defined by M and Ω which has values of $\widetilde{\Delta}_{\tau c} = \sup\limits_{0 \leq t \leq \infty} |\Delta_\tau(t)_c|$, which do not exceed K (0) $M\bar{\varepsilon}_{\tau\,c}$.

3. The majorants of $\sum_{k=2}^{\infty} |C_k|\Omega^k$. By definition, the distortion function

$$\varkappa_\tau(p) = \frac{1 + a_1 p + \cdots + a_m p^m}{1 + b_1 p + \cdots + b_n p^n} - e^{(a_1 - b_1)p}$$

is holomorphic within the circle $|p| < R$. Hence within this circle $\kappa_\tau(p)$ can be represented as a Maclaurin series:

$$\sum_{k=0}^{\infty} \frac{1}{k!} \varkappa_\tau^{(k)}(0)\, p^k = \sum_{k=2}^{\infty} C_k p^k.$$

The coefficients in this are functionally related to those in (23). In fact,

$$C_k = \frac{1}{k!} \varkappa_\tau^{(k)}(0) = \frac{1}{k!} \left[\varkappa_0^{(k)}(0) - \varkappa_0^{(1)k}(0)\right] = B_k - \frac{1}{k!} B_1^k.$$

Using this, together with (24a) and (24b) we can derive recurrence relations for the set of coefficients $\{C_k\}$:

$$C_k = \Delta_1 \left(A_{k-1} - \frac{1}{k!} \Delta_1^{k-1}\right) + \Delta_2 A_{k-2} + \Delta_3 A_{k+3} + \cdots, \qquad (38)$$

$$A_k = -b_1 A_{k-1} - b_2 A_{k-2} - b_3 A_{k-3} - \cdots, \qquad (39)$$

$$\Delta_i = a_i - b_i.$$

where

$$A_i = \begin{cases} 0 & \text{for} \quad i > 0, \\ 1 & \text{for} \quad i = 0, \end{cases} \qquad \begin{matrix} a_i = 0 & \text{for} & i > m, \\ b_i = 0 & \text{for} & i > n. \end{matrix}$$

(38) enables us to represent (37) as an infinite sum with terms defined by inequalities of the type

$$|C_2|\,\Omega^2 \leqslant |\Delta_1|\left|\left(A_1 - \frac{1}{2!}\Delta_1\right)\right|\Omega^2 + |\Delta_2|\,\Omega^2,$$

$$|C_3|\,\Omega^3 \leqslant |\Delta_1|\left|\left(A_2 - \frac{1}{3!}\Delta_1^2\right)\right||\Omega^3 + |\Delta_2|\,\Omega^3|\,A_1| + |\Delta_3|\,\Omega^3,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$|C_k|\,\Omega^k \leqslant |\Delta_1|\left|\left(A_{k-1} - \frac{1}{k!}\Delta_1^{k-1}\right)\right|\Omega^k + |\Delta_2||A_{k-2}|\,\Omega^k + \cdots + |\Delta_k|\,\Omega^k,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

This then gives

$$\epsilon_{\tau\,C} < \sum_{k=2}^{\infty}|C_k|\,\Omega^k \leqslant |\Delta_1| \sum_{k=2}^{\infty}\left|A_{k-1} - \frac{1}{k!}\Delta_1^{k-1}\right|\Omega^{k-i} + \\ + \sum_{i=2}^{l}|\Delta_i|\,\Omega^i\left(1 + \sum_{k=1}^{\infty}|A_k|\,\Omega^k\right).$$

Here l, as before, is the larger of m and n.

Remembering that

$$\sum_{k=1}^{\infty}|A_k|\,\Omega^k \leqslant \frac{\sum_{i=1}^{n}|b_i|\,\Omega^i}{1 - \sum_{i=1}^{n}|b_i|\,\Omega^i},$$

we can put $\epsilon_{\tau\,C}$ in the form

$$\epsilon_{\tau\,C} \leqslant \frac{\sum_{i=1}^{l}|\Delta_i|\,\Omega^i}{1 - \sum_{i=1}^{n}|b_i|\,\Omega^i} + e^{|\Delta_1|\Omega} - 1 - 2|\Delta_1|\,\Omega = \zeta_1 . \qquad (40)$$

We can use other majorants than the ξ_1 to estimate $\epsilon_{\tau\,C}$, i.e. the ones constructed by using the grouped parameters D_τ, Q_τ, etc., of Section 3.

The majorants, in particular, are

$$\xi_2 = \eta_2 + e^{|\Delta_1|\Omega} - 1 - 2|\Delta_1|\,\Omega,$$

$$\xi_3 = \frac{\Omega^2 D_\tau^2}{1 - \Omega D_\tau}$$

$$\xi_4 = e^{Q_\tau\Omega} - 1 - \Omega Q_\tau .$$

They are constructed on the same principles as the η_i.

Certain typical features of the estimates for $\epsilon_{\tau\,C}$ are worthy of note.

1. For all majorants $\{\xi_i\}$ $\lim\limits_{\Omega\to 0} \frac{d}{d\Omega}\,\xi_i = 0$, while $\lim\limits_{\Omega\to 0} \frac{d}{d\Omega}\,\eta_i \neq 0$. Hence for Ω sufficiently small $\xi_i < \eta_i$.

2. For all majorants $\{\xi_i\}$ $\lim\limits_{\Delta_i\to 0} \xi_i = \eta_i$. Hence for small values of

$$|\Delta_1| = \left|\frac{K^{(1)}(0)}{K(0)}\right| = |\varkappa_0^{(1)}(0)| = |a_1 - b_1|$$

the requirements as to precise followup and recording practically coincide.

These properties of the estimates for the $\epsilon_{\tau\,c}$ enable us to set practial bounds to the utility of this parameter for evaluating the reproduction properties of any particular system.

5. The Second Sufficient Condition for Precise Reproduction

1. Structure of $\alpha_{out}(t)_f$. We will derive the structure of the free component for the case where the transfer function and the class of inputs comply with the necessary conditions for precise reproduction. As before, we suppose that the transfer function and the transforms of all probable inputs are rational functions of p, all the poles of which lie in a finite region of the p plane.

Here z = 0, so from (11) we have

$$\alpha_{out}(t)_f = \sum_{k=1}^{N} \frac{1}{(s_k-1)!} \lim_{p\to-\beta_k} \frac{d^{s_k-1}}{dp^{s_k-1}} [A_{in}(p) K(p)(p+\beta_k)^{s_k} e^{pt}]. \tag{41}$$

Apply Leibnitz's formula to the components in the RHS in (41) and, passing to the limit, we can put

$$\alpha_{out}(t)_f = \sum_{k=1}^{N} e^{-\beta_k t} \sum_{i=0}^{s_k-1} \frac{1}{i!} t^i \sum_{j=0}^{s_k-1-i} \frac{1}{j!} A_{in}^{(j)}(-\beta_k) \frac{1}{(s_k-1-i-j)!} \times$$
$$\times \lim_{p\to-\beta_k} \frac{d^{s_k-1-i-j}}{dp^{s_k-1-i-j}} [K(p)(p+\beta_k)^{s_k}].$$

Since, in accordance with the notation previously adopted,

$$\frac{1}{(s_k-1-i-j)!} \lim_{p\to-\beta_k} \frac{d^{s_k-1-i-j}}{dp^{s_k-1-i-j}} [K(p)(p+\beta_k)^{s_k}] =$$
$$= K(0) L\varkappa_{s_1+\cdots+s_{k-1}+i+j},$$

then

$$\alpha_{out}(t)_f = K(0) L \sum_{k=1}^{N} e^{-\beta_k t} \sum_{i=0}^{s_k-1} \frac{1}{i!} t^i \times$$
$$\times \sum_{j=0}^{s_k-1-i} \frac{\varkappa_{s_1+s_2+\cdots+s_{k-1}+1+i+j}}{j!} A_{in}^{(j)}(-\beta_k). \tag{42}$$

When $\alpha_{in}(t)$ is reproduced the component $\alpha_{out}(t)_f$ occurs whenever $\alpha_{in}(t)$ or it derivatives $\alpha_{in}^{(k)}(t)$ show discontinuities.

In the general case an input with discontinuities at $t_1, t_2, \ldots, t_y$ can be represented as a sum of inputs:

$$\alpha_{in}(t) = \sum_{x=1}^{y} \alpha_{x\,in}(t-t_x),$$

where each component, plus its derivatives, is continuous for $t_x \leq t \leq \infty$ and precisely zero for $t < t_x$.

In this case the free motion will occur at $t = t_x$, so from (42) we have in each case:

when $t < t_x$

$$\Delta_x(t)_f = \alpha_{x\,out}(t)_f \equiv 0;$$

when $t \geq t_x$

$$\Delta_x(t)_f = \alpha_{x\ \text{out}}(t)_f = K(0)\, L \sum_{k=1}^{N} e^{-\beta_k(t-t_x)} \times$$

$$\times \sum_{i=0}^{s_k-1} \frac{1}{i!}(t-t_x)^i \sum_{j=0}^{s_k-1-i} \frac{\varkappa_{s_1+\cdots+s_{k-1}+1+i+j}}{j!} A_{\text{in}}^{(j)}(-\beta_k).$$

2. The second sufficient condition for precise reproduction. At the instants when $|\alpha_{\text{out}}(t)_f$ is comparable with $|\alpha_{\theta\text{out}}(t)|$ and $|\alpha_{\tau\text{out}}(t)|$ the system does not, strictly speaking, fulfill its function.

This is because we have required that the free motion be rapidly damped, which is imposed because of the necessary conditions for precise reproduction.

It is best to use T_f to evaluate the damping rate; it is given by

$$K(0)\, M\bar{\varepsilon}_f = \sup_{T_f < t < \infty} |\alpha_{\text{out}}(t)_f|.$$

If $T_f \leq \bar{T}_f$ and the first sufficient condition are complied with the system reproduces the set of inputs very well. In this case the metrological parameters $\varepsilon_{\theta C}$, $\varepsilon_{\tau C}$, and T_f will not exceed some preset values $\bar{\varepsilon}_{\theta C}$, $\bar{\varepsilon}_{\tau C}$, $\bar{T}_f$, respectively.

The second sufficient condition is thus that the inequality $T_f \leq \bar{T}_f$ be complied with.

3. The majorants of $\alpha_{\text{out}}^-(t)_f$. Using (42) with $A_{\text{in}}(p) = A_{\text{in}}^-(p) = \dfrac{M}{p}$ we get

$$\alpha_{\text{out}}^-(t)_f = -K(0)\, ML \sum_{k=1}^{N} \frac{e^{-\beta_k t}}{\beta_k} \sum_{i=0}^{s_k-1} \frac{1}{i!} t^i \sum_{j=0}^{s_k-1-i} \frac{\varkappa_{s_1+\cdots+s_{k+1}+1+i+j}}{\beta_k^{j+1}}.$$

Hence,

$$\bar{\varepsilon}_f = L \sup_{T_f < t < \infty} \left| \sum_{k=1}^{N} \frac{e^{-\beta_k t}}{\beta_k} \sum_{i=0}^{s_k-1} \frac{1}{i!} t^i \sum_{j=0}^{s_k-1-i} \frac{\varkappa_{s_1+\cdots+s_{k-1}+1+i+j}}{\beta_k^{j+1}} \right|.$$

As this expression is complicated it is very difficult to determine T_f. Let us consider ways of simplifying it.

Remembering that

$$\sup_{1<i<s_1+\cdots+s_N=n} |\varkappa_i| = 1, \qquad \sup_{1<k<N} s_k = s,$$

$$\sup_{1<k<N} \left|\frac{1}{\beta_k}\right| = R_b^{-1}, \qquad \sup_{1<k<N} \left|\frac{1}{a_k}\right| = a^{-1},$$

we may, from (42), write

$$|\alpha_{\text{out}}^-(t)_f| \leq K(0)\, LMN \frac{e^{-at}}{R_b^s} \frac{R_b^s - 1}{R_b - 1} \left\{ 1 + \frac{1}{1!} \frac{R_b^{s-1} - 1}{R_b^s - 1} R_b t + \cdots \right.$$

$$\left. \cdots + \frac{1}{(s-1)!} \frac{R_b - 1}{R_b^s - 1} R_b^{s-1} t^{s-1} \right\} = F_1(t).$$

The majorant $F_1(t)$ is correct for any arrangement of the N zeros of b(p) of degree $n \leq N_1 s$ in the semiplane $\text{Re}(p) \leq -a$ (excluding the area within the circle $|p| = R_b$).

This favorably differentiates $F_1(t)$ from $F(t)$;

$$F(t) = K(0)\,Me^{-at}\left\{1+\frac{1}{1!}at+\cdots+\frac{1}{(n-1)!}a^{n-1}t^{n-1}\right\} \geqslant |\alpha_{out}^{-}(t)_f|,$$

proposed by Fel'dbaum for certain classes of distribution for the zeros of $b(p)$ of degree n in the semiplane $Re(p) \leq -a$ [5].

However, the majorants of greatest practical interest are

$$F_2(t) = K(0)\,LMN\frac{R_b^s-1}{R_b-1}\,\frac{e^{-(a-1)t}}{R_b^s} = K(0)\,MF_2e^{-(a-1)t} \quad \text{for} \quad a>1$$

and

$$F_3(t) = K(0)\,LMN\frac{(R_b^s-1)t^{-1}}{aeR_b^s(R_b-1)}\left\{1+\frac{2^2}{1!}\frac{R_b^{s-1}-1}{R_b^s-1}\frac{R_b}{ae}+\cdots+\frac{S^s}{(s-1)!}\frac{R_b-1}{R_b^s-1}\left(\frac{R_b}{ae}\right)^{s-1}\right\} =$$
$$= K(0)\,MF_3t^{-1} \quad \text{for} \quad a>0,$$

which are suitable for finding T_f rapidly.

The first is constructed using

$$1+\frac{1}{1!}\frac{R_b^{s-1}-1}{R_b^s-1}R_bt+\frac{1}{2!}\frac{R_b^{s-2}-1}{R_b^s-1}R_b^2t^2+\cdots+\frac{1}{(s-1)!}\frac{R_b-1}{R_b^s-1}R_b^{s-1}t^{s-1}<e^t,$$

which is correct for $R_b > 1$ and $t > 0$; the second by using

$$t^ie^{-at} \leqslant \left[\frac{i+1}{ae}\right]^{i+1}t^{-1} \quad \text{for} \quad t>0.$$

Substituting

$$t^ie^{-at} \leqslant \left[\frac{i}{ae}\right]^i \text{for} \quad t>0$$

into $F_1(t)$ we also get an estimate $M_f = \sup\limits_{0 \leq t \leq \infty} \alpha_{out}(t)_f$:

$$M_f \leqslant \frac{K(0)\,LMN}{R_b^s(R_b-1)}\left\{1+\frac{1}{1!}\frac{R_b^{s-1}-1}{R_b^s-1}\,\frac{R_b}{ae}+\cdots+\frac{(s-1)^{s-1}}{(s-1)!}\,\frac{R_b-1}{R_b^s-1}\left(\frac{R_b}{ae}\right)^{s-1}\right\} = F_4.$$

Hence $F_1(t)$, $F_2(t)$ and $F_3(t)$ enable us to show the second sufficient condition either as

$$\frac{1}{a-1}\ln\frac{F_2}{\varepsilon_f} \leqslant \overline{T}_f \qquad \text{for} \quad a>1,$$

or as

$$\frac{F_3}{\varepsilon_f} \leqslant \overline{T}_f \qquad \text{for} \quad a>0.$$

We finally observe that $F_1(t)$, $F_2(t)$, and $F_3(t)$ remain the majorants $|\alpha_{out}^{-}(t)_f|$ when R_b is replaced by a. If the coefficients in $b(p)$ are always positive we can also replace R_b by the lower bounds to the zeros of this polynomial:

$$\lambda_b = \left[\sup_{0<i<n}\frac{b_i}{b_{i-1}}\right]^{-1}.$$

SUMMARY

1. The methods given for evaluating the reproducing properties of linear systems enable one to decide on the correspondence between the system and the class of inputs which will ensure a given precision (in followup or recording).

2. These methods may be of value in solving problems of analysis, synthesis and correction for measurement, servo and control systems.

LITERATURE CITED

[1] Ia. Z. Tsypkin, Fundamentals of the Theory and Calculation of Reproducing Systems [in Russian], Cand. Diss. (1945).

[2] S. P. Strelkov, "The general theory of linear amplifiers.I," Automation and Remote Control(USSR) 9, No. 3 (1948).

[3] S. P. Strelkov, "The general theory of linear amplifiers.II," Automation and Remote Control (USSR) 10, No.4 (1949).

[4] A. A. Fel'dbaum, "The distributions of the roots of the characteristic equations of automatic control systems," Automation and Remote Control (USSR) 9, No. 4 (1948).

[5] A. A. Kharkevich, "Use of the mean-square error in evaluating mechanical converters," J. Exptl. Theoret Phys. (USSR) No. 7 (1937).

[6] V. V. Solodovnikov, "The dynamic accuracy and optimal characteristics of systems and conversion devices, " Dokl. Akad. Nauk. SSSR 27, No.2 (1951).

[7] V. V. Solodovnikov, Introduction to the Statistical Dynamics of Automatic Control Systems [in Russian] (1952).

[8] H. James, N. Nichols and R. Phillips, Theory of Servomechanisms [Russian translation] (IL, 1948).

[9] I. I. Privalov, Introduction to the Theory of Functions [in Russian] (1948).

[10] Walle-Poussen, Course in Infinitesimal Analysis (1953).

[11] A. G. Kurosh, Course in Higher Algebra [in Russian] (1955).

[12] Diens, The Taylor Series,(Oxford, 1911).

Received May 13, 1957

THE PROBLEM OF SYNTHESIZING LINEAR DYNAMIC SYSTEMS WITH VARIABLE PARAMETERS

A. M. Batkov
(Moscow)

A method is presented for determining the differential equation of a linear dynamic system with variable parameters, when the impulse transient response is given.

One of the important problems in the synthesis of automatic control systems is the determination of a design to implement a known impulse transient response.

In the case of systems whose processes are described by linear differential equations with constant coefficients, this problem is solved by determining the transfer function corresponding to the known impulse transient response of the system, and then representing this transfer function in the form of a product of elementary factors.

If the system is described by a linear differential equation with variable coefficients, then its transfer function depends not only on frequency, but also on time [1], and this, in the majority of cases, renders it difficult to determine the design for implementing the system. Because of this, one usually proposes to determine the design of systems with variable parameters by starting out from the differential equation characterizing the system. Below, we consider the problem of determining this differential equation in the cases where the impulse transient response is given.* The questions involved in determining the design corresponding to the equation are considered in the examples.

Let there be given an impulse transient response $W(t, \tau)$ of a linear dynamic system with variable parameters. We assume that the differential equation for this system has the form

$$L(p, t)\,u(t) = M(p, t)\,v(t), \tag{1}$$

where the operators $L(p, t)$ and $M(p, t)$ are defined by the expressions

$$L(p, t) = \sum_{i=0}^{n} a_i(t)\,p^{n-i}, \qquad M(p, t) = \sum_{j=0}^{m} b_j(t)\,p^{m-j}, \tag{2}$$

and $p = d/dt$, $u(t)$ is the output and $v(t)$ the arbitrary input variables of the system, the variable coefficients, $a_i(t)$ and $b_j(t)$, are continuous functions of time, and $\underline{n}$ and $\underline{m}$ are unknown constant integers.

One of the possible methods of determining these quantities is expounded below.

It is known [2] that the output variable of System (1) may be written in the form

$$u(t) = \int_0^t W(t, \tau)\,v(\tau)\,d\tau, \tag{3}$$

* This happens, for example, in problems of synthesizing optimum filters.

where W(t, τ) is the impulse transient response of the system, and is assumed to be known.

From Expression (3) it is easy to determine the initial value of u(t) and its derivatives:

$$
\begin{aligned}
&u(t)|_{t=0} = 0,\\
&u'(t)|_{t=0} = w(0)\,v(0),\\
&u''(t)|_{t=0} = 2w'(0)\,v(0) + w(0)\,v'(0),\\
&\dots\dots\dots\dots\dots\dots\\
&u^{(k)}(t)|_{t=0} = kw^{(k-1)}(0)\,v(0) + \frac{k(k-1)}{1:2}\,w^{(k-2)}(0)\,v'(0) + \cdots\\
&\qquad + kw'(0)\,v^{(k-2)}(0) + w(0)\,v^{(k-1)}(0).
\end{aligned}
\tag{4}
$$

The problem consists of determining, in the form of (3), the solution of differential Equation (1), corresponding to the initial Conditions (4).

Writing the general solution of the homogeneous equation

$$L(p, t)\,u(t) = 0 \tag{5}$$

in the form

$$u(t) = c_1(t)\,u_1(t) + c_2(t)\,u_2(t) + \cdots + c_n(t)\,u_n(t) \tag{6}$$

and applying the method of varying the arbitrary constants [3], we obtain a non-homogeneous system of n equations in the derivatives, $c_1'(t), \ldots, c_n'(t)$:

$$
\begin{aligned}
&u_1c_1' + u_2c_2' + \ldots + u_nc_n' = 0,\\
&u_1'c_1' + u_2'c_2' + \ldots + u_n'c_n' = 0,\\
&\dots\dots\dots\dots\dots\dots\\
&u_1^{(n-1)}c_1' + u_2^{(n-1)}c_2' + \ldots + u_n^{(n-1)}c_n' = M(p, t)\,v(t).
\end{aligned}
\tag{7}
$$

Since the determinant $\Delta(t)$ of this system is the Wronskian of Equation (5), it is different from zero and, consequently, System (7) has a unique solution, which may be written in the form

$$c_i'(t) = \frac{\partial \ln \Delta(t)}{\partial u_i^{(n-1)}(t)}\,M(p, t)\,v(t) \quad (i = 1, 2, \ldots, n).$$

Hence,

$$c_i(t) = \int \frac{\partial \ln \Delta(t)}{\partial u_i^{(n-1)}(t)}\,M(p, t)\,v(t)\,dt + \gamma_i. \tag{8}$$

The constants γ_i are determined from the initial Conditions (4). In order to so determine them, we substitute u(t) and its first n-1 derivatives from (6) into (4). Since the non-homogeneous system of equations thus obtained has its determinant equal to $\Delta(0)$, it, as with (7), has a unique solution

$$c_i(0) = \frac{\begin{vmatrix} u_1(0) \ldots u_{i-1}(0) & 0 & u_{i+1}(0) \ldots u_n(0)\\ u_1'(0) \ldots u_{i-1}'(0) & w(0)\,v(0) & u_{i+1}'(0) \ldots u_n'(0)\\ \dots & \dots & \dots \end{vmatrix}}{\Delta(0)} = \frac{\Delta_i^*(0)}{\Delta(0)} \; (i = 1, 2, \ldots, n). \tag{9}$$

Since, from (8),

$$\gamma_i = c_i(0) - \left[\int \frac{\partial \ln \Delta(t)}{\partial u_i^{(n-1)}(t)}\,M(p, t)\,v(t)\,dt\right]_{t=0},$$

then, comparing (8) and (9), we obtain finally,

$$c_i(t) = \int_0^t \frac{\partial \ln \Delta(\tau)}{\partial u_i^{(n-1)}(\tau)} M(p, \tau) v(\tau) d\tau + \frac{\Delta_i^*(0)}{\Delta(0)} \quad (i = 1, 2, \ldots, n). \tag{10}$$

Consequently, a particular solution of the linear, non-homogeneous differential Equation (2) with variable coefficients and initial Conditions (4) may be written in the form

$$u(t) = \sum_{i=1}^{n} \left[\int_0^t \frac{\partial \ln \Delta(\tau)}{\partial u_i^{(n-1)}(\tau)} M(p, \tau) v(\tau) d\tau + \frac{\Delta_i^*(0)}{\Delta(0)}\right] u_i(t), \tag{11}$$

where $u_1(t), \ldots, u_n(t)$ are particular solutions of the homogeneous Eqautions (5).

Equating u(t) from (3) and (11), we obtain relationships for determining the particular solutions, $u_1, u_2, \ldots$ $\ldots, u_n$, and the right member, $M(p, \tau)$, of the desired differential equation in the form

$$\int_0^t \left[W(t, \tau) v(\tau) - \sum_{i=1}^{n} u_i(t) \frac{\partial \ln \Delta(\tau)}{\partial u_i^{(n-1)}(\tau)} M(p, \tau) v(\tau)\right] d\tau = \sum_{i=1}^{n} u_i(t) \frac{\Delta_i^*(0)}{\Delta(0)}. \tag{12}$$

Since the systems considered in this article are described by linear differential equations with variable coefficients then, as a consequence of the theorem as to the existence of solutions of these equations [3], their impulse transient responses may always be given in the form*

$$W(t, \tau) = \sum_{i=1}^{n} \varphi_i(t) \psi_i(\tau), \tag{13}$$

where the $\varphi_i(t)$ are linearly independent functions, constituting a basis for the homogeneous differential Equations (5), corresponding to (13), and n is the order of the equation.

If, in (12), we make the substitution, $\varphi_i(t) \equiv u_i(t)$ $(i = 1, 2, \ldots, n)$, we obtain the basis for the homogeneous differential equation, which coincides, to within a factor independent of the derivatives, with the left member of the desired equation.

Employing well-known methods [3], we determine the left member of the equation by expanding the (n + 1)-order determinant:

$$D(t) = \begin{vmatrix} u_1 u_2 \ldots u_n u \\ u_1' u_2' \ldots u_n' u' \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot \\ u_1^{(n)} u_2^{(n)} \ldots u_n^{(n)} u^{(n)} \end{vmatrix}. \tag{14}$$

With Condition (13), and as a consequence of the linear independence of the functions $u_i(t)$ $(i = 1, 2, \ldots, n)$, the relationships (12) can be written as a system of n equations of the form

$$\int_0^t \left[\frac{\partial \ln \Delta(\tau)}{\partial u_i^{(n-1)}(\tau)} M(p, \tau) v(\tau) - \psi_i(\tau) v(\tau)\right] d\tau + \frac{\Delta_i^*(0)}{\Delta(0)} = 0, \tag{15}$$

from which the form of the left member of the equation, i.e., M(p, t), is determined.

---

* We do not consider here the questions regarding transforming the given function $W(t, \tau)$ to the form (13), and the corresponding identity of the impulse transient response of the system and its differential equation.

The computations are significantly simplified if it is necessary to determine a system of n linear differential equations of the first order, with variable coefficients, corresponding to the given impulse transient response of Form (13).

In fact, taking $W(t, \tau)$ as the sum of n functions, $W_i(t, \tau) = \varphi_i(t)\psi_i(\tau)$, in accordance with the reasoning above we obtain a system of n first-order differential equations, the operators of which are defined by expressions

$$L_i(p, t)\, u(t) = \frac{du(t)}{dt} - \frac{1}{\varphi_i(t)} \frac{d\varphi_i(t)}{dt} u(t),$$

$$\psi_i(t) = M_i^*(p, t) \frac{1}{\varphi_i(t)} \quad (i = 1, 2, \ldots, n),$$

where $M_i^*(p, t)$ is the operator conjugate to $M_i(p, t)$.

It is assumed that these equations can be physically realized, i.e., that the system with the impulse transient response (13) can be implemented by n linear, first-order dynamic systems, connected in parallel. This is a very convenient approach from the point of view of applying linear analogue devices.

The proposed method is completely applicable to systems with constant parameters. The method permits one to determine the form of the differential equation and, consequently, the design of the system in the cases when the functions entering into $W(t, \tau)$ are given in general form. It is clear that in such cases, even when the system has constant parameters, the method of determining the structure and parameters of the system from the transfer function is inapplicable. An example corresponding to this case is considered below.

Simple examples to illustrate the application of the method.

Example 1. In [1] there is considered a system, the impulse transient response of which is equal to

$$W(t, \tau) = (1 + \rho \cos \omega_0 \tau) \exp\left\{-\left[(t - \tau) + \frac{\rho}{\omega_0} (\sin \omega_0 t - \sin \omega_0 t)\right]\right\}.$$

In this case, $n = 1$, and $\varphi_1(t) = \exp\left\{-\left(t + \frac{\rho}{\omega_0} \sin \omega_0 t\right)\right\} = \Delta(t)$.

The determinant corresponding to (14) will, in this case, have the form:

$$D(t) = \begin{vmatrix} \exp\left\{-\left(t + \frac{\rho}{\omega_0} \sin \omega_0 t\right)\right\} & u \\ -\exp\left\{-\left(t + \frac{\rho}{\omega_0} \sin \omega_0 t\right)\right\}(1 + \rho \cos \omega_0 t) & u' \end{vmatrix}.$$

Expanding the determinant, and discarding the multiplier $\exp\left\{-\left(t + \frac{\rho}{\omega_0} \sin \omega_0 t\right)\right\}$, we obtain the left member of the desired differential equation:

$$L(p, t)\, u(t) = u' + (1 + \rho \cos \omega_0 t)\, u.$$

For the determination of the right member, we find $\Delta(0) = 1$, $\Delta_1'(0) = 0$. Equation (15) then takes the form

$$\int_0^t \exp\left\{\left(\tau + \frac{\rho}{\omega_0} \sin \omega_0 \tau\right)\right\} [M(p, \tau)\, v(\tau) - (1 + \rho \cos \omega_0 \tau)\, v(\tau)]\, d\tau = 0.$$

$$M(p, \tau) = 1 + \rho \cos \omega_0 \tau.$$

Thus, we obtain the equation for the system:

$$\frac{du}{dt} + (1 + \rho \cos \omega_0 t)\, u = (1 + \rho \cos \omega_0 t)\, v(t).$$

From this it is clear that we have a first-order system which can be realized by an RC filter, with variable resistance

$$R(t)=\frac{1}{1+\rho\cos\omega_0 t}\text{ and }C=1.$$

As determined in [1], the approximate transfer function for $\rho \ll 1$ will be

$$H(j\omega, t)=\frac{1}{1+j\omega}+\rho\cos\omega_0 t\frac{j\omega}{(1+j\omega)^2+\omega_0^2}+$$
$$+\rho\omega_0\sin\omega_0 t\frac{j\omega}{[(1+j\omega)^2+\omega_0^2](1+j\omega)}+O(\rho^2).$$

Whence the impossibility of drawing any conclusion as to the design of the system.

Example 2. It was recently shown [4] that for optimum filtering of a random signal in the presence of white noise, the linear system with variable parameters would have an impulse transient response of the form

$$W(t,\tau)=\frac{g(t)\,g(\tau)}{N^2+\int\limits_0^t g^2(\theta)\,d\theta},$$

where $g(t)$ is a known function.

We shall determine, by the method expounded in this article, system design and parameters corresponding to $W(t, \tau)$.

Obviously,

$$n=1\text{ and }\varphi_1(t)=\frac{g(t)}{N^2+\int\limits_0^t g^2(\theta)\,d\theta},$$

$$D(t)=\begin{vmatrix}\dfrac{g(t)}{N^2+\int\limits_0^t g^2(\theta)\,d\theta}, & u\\ \dfrac{g'(t)\left[N^2+\int\limits_0^t g^2(\theta)\,d\theta\right]-g^3(t)}{\left[N^2+\int\limits_0^t g^2(\theta)\,d\theta\right]^2}, & u'\end{vmatrix}.$$

This gives us the left member of the equation

$$L(p,t)\,u(t)=u'-\frac{g'(t)\left[N^2+\int\limits_0^t g^2(\theta)\,d\theta\right]-g^3(t)}{\left(N^2+\int\limits_0^t g^2(\theta)\,d\theta\right)g(t)}\,u.$$

For determining $M(p, t)$, we employ (15), obtaining the equation

$$\int\limits_0^t\left[\frac{N^2+\int\limits_0^\tau g^2(\theta)\,d\theta}{g(\theta)}M(p,\tau)\,v(\tau)-g(\tau)\,v(\tau)\right]d\tau=0.$$

From this

$$M(p, t) = \frac{g^2(t)}{N^2 + \int_0^t g^2(\theta)\, d\theta}.$$

Consequently, the system under consideration is described by the differential equation

$$\frac{du}{dt} - \left[\frac{g'(t)}{g(t)} - \frac{g^2(t)}{N^2 + \int_0^t g^2(\theta)\, d\theta}\right] u = \frac{g^2(t)}{N^2 + \int_0^t g^2(\theta)\, d\theta} v(t)$$

and may thus be realized, for example, by a first-order RC filter, in which R and C vary with time according to

$$C(t) = \frac{C_0}{g(t)}, \quad R(t) = \frac{N^2 + \int_0^t g^2(\theta)\, d\theta}{C_0 g(t)}.$$

Example 3. We consider now an application of our method to the determination of the differential equation for a system with constant parameters.

Let

$$W(t, \tau) = e^{-a(t-\tau)} \cos \omega_0 (t - \tau). \tag{16}$$

Transforming $\cos \omega_0(t - \tau)$ by well-known trigonometric formulas, we obtain $n = 2$, $\varphi_1(t) = e^{-at} \cos \omega_0 t$, $\varphi_2(t) = e^{-at} \sin \omega_0 t$. Setting up the determinant D(t) according to Formula (14), we find that

$$L(p, t) = (p + a)^2 + \omega_0^2.$$

Since $\Delta(t) = \omega_0 e^{-2at}$,

$$\frac{\partial \ln \Delta(t)}{\partial \varphi_1'(t)} = -\frac{1}{\omega_0} e^{at} \sin \omega_0 t, \quad \frac{\partial \ln \Delta(t)}{\partial \varphi_2'(t)} = \frac{1}{\omega_0} e^{at} \cos \omega_0 t,$$

we get the following system of equations for determining the right member of the equation:

$$\int_0^t e^{a\tau} \left[\cos \omega_0 \tau v(\tau) + \frac{1}{\omega_0} \sin \omega_0 \tau M(p, \tau) v(\tau)\right] d\tau = 0,$$

$$\int_0^t e^{a\tau} \left[v(\tau) \sin \omega_0 \tau - \frac{1}{\omega_0} \cos \omega_0 \tau M(p, \tau)\, v(\tau)\right] d\tau = 0.$$

Integrating by parts and adding, we find

$$\int_0^t e^{a\tau} \sin \omega_0 (t - \tau) \left[M(p, \tau) v(\tau) - v'(\tau) - av(\tau)\right] d\tau = 0.$$

Thus $M(p, t) = p + a$ and the desired differential equation has the form:

$$\frac{d^2u}{dt^2} + 2a \frac{du}{dt} + (a^2 + \omega_0^2)\, u = \frac{dv}{dt} + av.$$

It is easy to verify that the direct application of the Laplace transform to the impulse transient response, (16), leads to an analogous result.

LITERATURE CITED

[1] L. A. Zadeh, "Frequency analysis of variable networks", Trans. IRE 38, 3 (March 1950).

[2] K. S. Miller, Properties of Impulsive Responses and Green's Functions. IRE Transactions on circuit theory, vol. CT-2, (1955).

[3] V. V. Stepanov, Course in Ordinary Differential Equations [in Russian] (Gostekhizdat, 1950).

[4] V. V. Solodovnikov and A. M Batkov, A method of determining the optimal characteristics of a class of self-adjusting systems, Automation and Remote Control (USSR) 18, 5 (1957).

Received October 27, 1957

TELEMETERING SYSTEMS WITH PULSE-CODE MODULATION

G. V. Burdenkov
(Moscow)

High-speed pulse-code telemetering devices are investigated. The possibility is demonstrated of constructing telemetering deivces with magnetic elements having rectangular hysteresis loops, together with crystal triodes and diodes. The basic parameters of such devices are cited, and an estimate is given of the accuracy of telemetering.

Experiments in the use of telemetering devices in power systems show that the unstable functioning of these telemetering devices is to a large measure conditioned by the influence of interference in the channel, interference which is particularly noticeable when the channel is complex or extensive. Hence the significance of questions of stability under interference of telemetering systems, and of questions of the independence of telemetered findings from variations in the parameters of the channel.

As is well known, among the telemetering systems most stable under interference is the pulse-code system, the functioning of which is the most independent of the channel's state, since the discrete method of telemetering does not entail an accurate reproduction of pulse form, but requires only a determination of the presence or absence of a pulse (see Fig. 1).

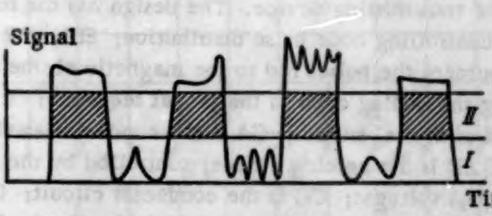


Fig. 1. Recovery of the code on the receiving end.
I) Clipped below; II) Clipped above.
Shadings mark the recovered pulses.

The very simplest principle of code formation is that of numerical coding, i.e., the construction of a sequence of combinations using one of the standard number systems, e.g., decimal or binary. Binary coding is usually employed in high-speed pulse-code systems, since the binary code, for the same accuracy in measurement, advantageously distinguishes itself from the decimal code by significant decrease in the necessary number of pulses, thus permitting a desirable decrease in the magnitude of error in discreteness for sufficiently high speeds of functioning. The complexity of implementing the coding and decoding must be mentioned as a disadvantage of these methods. The most well-known forms of such coding are:

a) Pulse-count coding;
b) Inverted pulse-count coding;
c) Coding by the principle of comparison with reading pulses;
d) Coding by means of variations in the accumulator voltage;
e) Coding using a comparison scheme with a multi-step comparison voltage [1 - 4].

Transmission may thus be carried out either by qualitatively different pulses, or by the presence or absence of homogeneous pulses.

Not all known forms of coding satisfy the requirements stemming from the conditions under which telemetering must be employed in power systems (as to accuracy, stability, etc.).

From this point of view, there is a good deal of interest in methods of coding using comparison schemes with multi-step comparison voltages (for brevity, these will be called "methods of multi-step equalization"), which were taken as the basis of development of pulse-code telemetering devices.

Particular attention was paid to the choice of elements for the apparatus, including the use of magnetic elements with rectangular hysteresis loops in conjunction with crystal triodes and diodes, with respect to the variation of the element parameters which, in the given case, has little influence on the accuracy of the telemetering, since we are dealing here with bi-stable devices, analogous to relays. From this was envisioned the possiblity of modular devices, the number of modules, or cells, to depend on the required number of elements in the code, as determined from the inequality, $\delta_D \leq \pm[2(2^n - 1)]^{-1}$, where δ_D is the discreteness error and $\underline{n}$ is the number of elements (pulses) in the code.

The frequency F with which code pulses may succeed one another is a function of the channel band-width, Δf.

For our method of telemetering coding, as already mentioned, it suffices merely to distinguish the presence or absence of a pulse. Therefore, it is possible to limit the band-width, $\Delta f = 2F$.

Code transmission time (time for one cycle) is found from the formula

$$T_c = n(t_u + t_p) + t_d + t_m,$$

where $\underline{n}$ is the number of elements in the code, t_u is the length of one code pulse, t_p is the length of the pause between code pulses, t_d is the length of the pause between code groups, and t_m is the length of the synchronizing pulse (the marker).

In the device developed at the Central Laboratories for Scientific Research in Electrical Engineering (TsNIEL), the choice of parameters was: $n = 6$, $F = 40$ c/sec $t_u = t_p = t_d$, $t_m = 3t_u$. The time for one cycle, $T_c = 0.2$ sec.

Fig. 2 is a block schematic of the transmitting device. The design has the following component parts: MV, a multivibrator, generating pulses for controlling code pulse distribution; EF1, EF2 and EF3, emitter followers for load-matching; D is a two-way distributor of the pulses fed to the magnetic elements with rectangular hysteresis loops, I, I', etc., these pulses switching the coding cells in the proper sequence; T is a trigger assembly, for switching the conductor assembly, comprised of crystal triodes; CA is the conductor assembly (in this circuit, electronic switching is performed by 6H2 tubes); ZD is the zeroing device, controlled by the sign of the difference between the measured (U_1) and the equalizing (U_2) voltages; CC is the conductor circuit; CB1 is the coincidence block, generating pulses for cutting off the corresponding trigger cell; CB2 is the block which isolates the output code pulses; SC1 and SC2 are shaping cascades for transforming the input pulses to rectangular pulses.

Coding is accomplished in the following manner:

As a preliminary step, the measured quantity, varying with time, is transformed to a rectified voltage, U_1. During the coding process, this voltage is compared with the equalizing voltage, U_2. This latter voltage varies stepwise, the amplitude of the steps being proportional to the decreasing binary series, $2^{n-1}, 2^{n-2}, 2^{n-3}, \ldots, 2^0$. The zeroing device is controlled by the sign of the voltage difference, $U_1 - U_2$. If this difference is positive, then the corresponding binary step is retained. For a negative difference, the corresponding binary step is not utilized in the formation of the equalizing voltage. In order to retain a positive difference while simultaneously switching to the next binary step, a pulse is sent through the path of CB2, this pulse corresponding to the previous term of the code.

Let us clarify the coding process by way of an example (Fig. 3). Assuming a six-valued code is being used, if the measured voltage $U_1 = 45U_0$ (U_0 being taken as the relative unit), then the following steps must be included in the equalizing voltage: 2^5U_0, 2^3U_0, 2^2U_0 and 2^0U_0. This will give the equalizing voltage $U_2 = (32 + 8 + 4 + 1)U_0 = 45U_0 = U_1$. Thus, code pulses 2^5, 2^3, 2^2, and 2^0 will be transmitted from time 2 through time 13 in the channel. Each cycle is preceded by a synchronizing pulse (marker) whose length is significantly different from that of the

code pulses. The marker is produced by element M of the two-way distributor and by shaper SC2. Distributor elements CT and BP prepare the circuit for the following code cycle. One of the basic components of the transmitting device is the two-way magnetic distributor, working with elements having rectangular hysteresis loops [5].

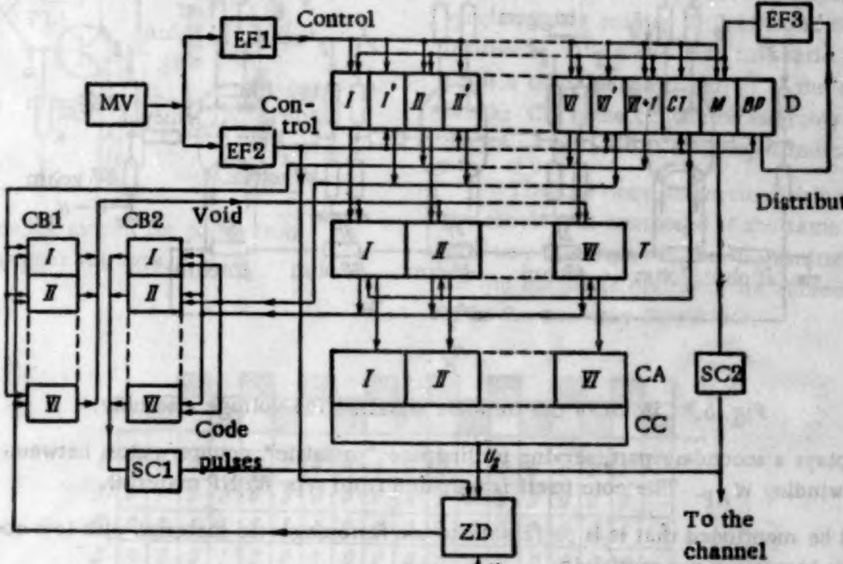


Fig. 2.

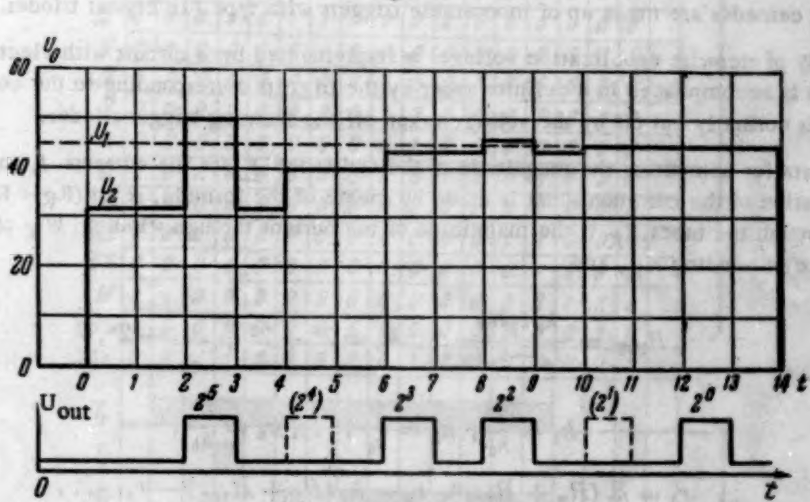


Fig. 3. Graph of the coding process.

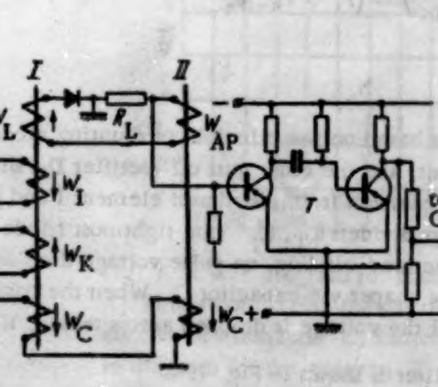


Fig. 4. Circuit of the zeroing device.

The comparison element in the transmitter design is the zeroing device (Fig. 4). Its design is comprised of two magnetic elements, I and II, with rectangular hysteresis loops, and a trigger T with one stable state.

A current proportional to voltage U_1 flows through winding W_I and a current, from the conductor circuit, proportional to compensating voltage U_2, flows through winding W_K. Depending on the sign of the voltage difference $U_1 - U_2$, core I is or is not primed to be switched.

If the compensating voltage is the larger of the two, then core I is primed and, when sensed via control winding W_C, causes a pulse to be generated at winding W_L.

which in turn transfers trigger T. This then provides a gating voltage to coincidence block CB1 and yields a pulse to switch off the memory trigger corresponding to the proper step of compensating (equalizing) voltage.

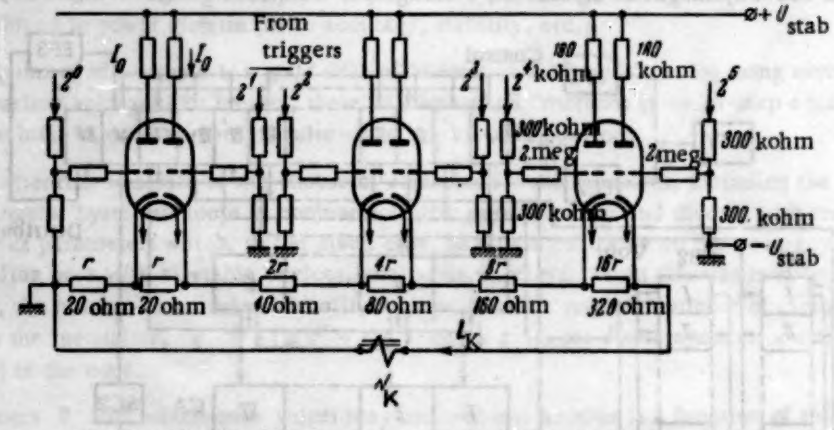


Fig. 5. Circuit of the stepwise equalization voltage assembly.

Core II plays a secondary part, serving to dissipate "parasitic" counter pulses between load winding W_L and anti-parasitic winding W_{AP}. The core itself is prepared from type 65 NP material.

It should be mentioned that it is preferable to use ferromagnetic material with low coercivity for the zeroing device, thus increasing the sensitivity.

The shaping cascades are made up of monostable triggers with type P1B crystal triodes.

The assembly of stepwise equalization voltages is implemented by a circuit with electronic switches, the switching of which is accomplished in a definite order by the triggers corresponding to the code cells (Fig. 5). The electronic switch is normally cut off by the voltage taken off the shunting trigger triodes.

The basic data for computing the magnitude of the resistance r are the currents I_0 and I_K and the resistance R_K. Calculation of the common shunt is made by means of the formula, $r = k(R_K + R_{tot})$, where I_0 is the current flowing through the tubes, I_K is the magnitude of the current through winding W_K of the zeroing device, R_K is the resistance of winding W_K, and

$$R_{tot} = \frac{k_n I_K R_K}{I_0 - k_n I_K}, \quad k_0 = 2^0 + 2^1 + \ldots + 2^n,$$

$$k_1 = \frac{2^0}{k_0}, \; k_2 = \frac{2^1}{k_0}, \ldots, k_n = \frac{2^n}{k_0},$$

$$r_1 = k_1 (R_K + R_{tot}), \; r_2 = k_2 (R_K + R_{tot}) - r_1, \ldots,$$

$$r_n = k_n (R_K + R_{tot}) - (r_1 + r_{n-1}).$$

Here, $k_0, \ldots, k_n$ are coefficients.

The coincidence circuit (Fig. 6) of the transmitter is based on the principle of shunting diode switches. With such a circuit, one of the compared voltages feeds the input, and the other cuts off rectifier D_2, shunting the output. The circuit is controlled by the coincidence of pulses taken from distributor element I and gating voltages from trigger T. With this voltage present across the voltage dividers R_T, R_T' (the rightmost triode of the trigger being cut off), diode D_2 does not pass current, and, with the load winding, no pulse voltage drop occurs across resistor R_D; the pulse is transmitted without loss to the pulse shaper via capacitor C. When the trigger reverses, the gating voltage is absent at diode D_2, and a large portion of the voltage is dropped across resistor R_D ($R_D \gg R_L$).

The interaction of the sub-assemblies of the transmitter is shown in Fig. 7.

Fig. 8 is a block schematic of the receiver. The design consists of the following blocks: A is an amplifier;

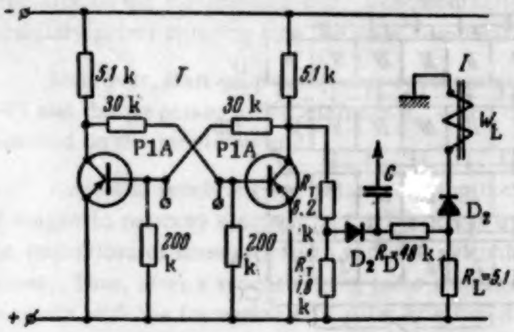


Fig. 6. Coincidence circuit for pulses from element I and trigger voltages.

AC is an amplifier-clipper, used for two-sided clipping of the channel pulses; MV is a multivibrator which generates pulses to control the magnetic distributor; TS is the time selector, used to discriminate the longer synchronizing pulses; EF1, EF2 and EF3 are emitter followers; D is a one-way magnetic pulse distributor; CB is a coincidence block; T is the memory trigger block; CA is the conductor assembly; CC is the conductor circuit; OI is the output indicating unit.

The one-way magnetic distributor employed in the receiver is composed of the same elements as the two-way distributor of the transmitter. The parameters for the magnetic elements are chosen to be the same as in the two-way distributor.

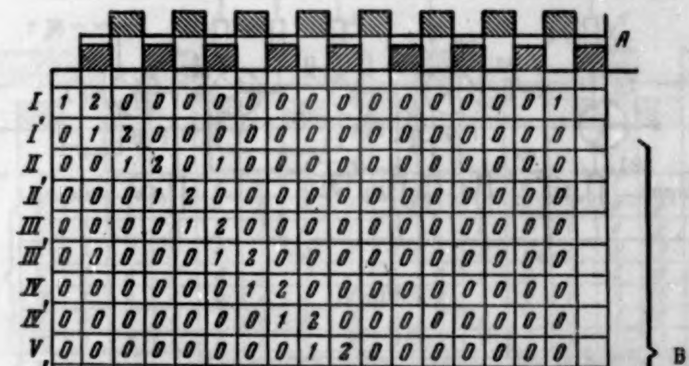


Fig. 7. Cycle diagram for transmitter functioning.
A) Pulses controlling the distributor; B) Switching diagram of the distributor elements; C) Switching of the zeroing device; D) Switch-in diagram for the triggers; E) Pulses entering the shaping cascades; F) Pulses in the channel. 1) element is primed; 2) moment of switching; 0) element "void."

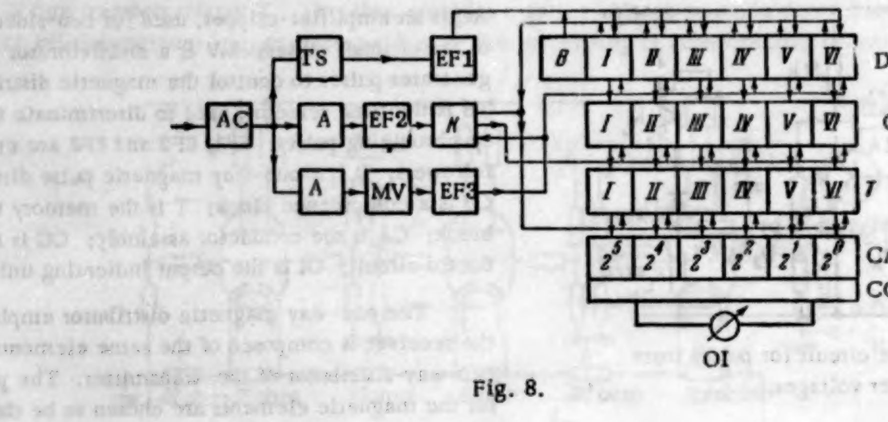


Fig. 8.

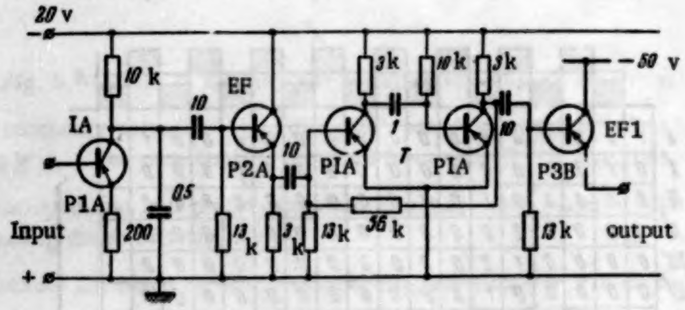


Fig. 9.

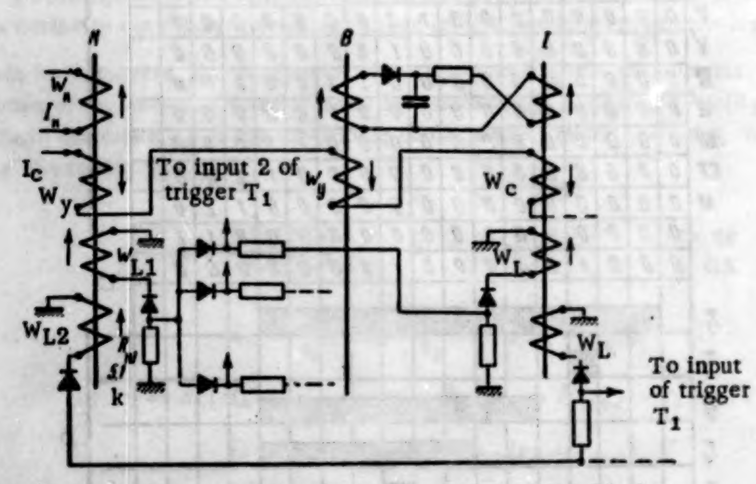


Fig. 10. Interaction of memory elements K with the distributor elements.

The pulse time selector (Fig. 9) consists of an intermediate amplifier IA, a trigger with one stable state, T, and emitter follower EF. The selection principle is based on the selection of a definite time constant at the input of the intermediate amplifier. Such a selector is frequently called an integrator. If a short pulse of length τ_s and a long pulse of length τ_l appear alternately at the input to the selector, then at the output of the amplifier we obtain the ratio of the peak voltages, which is approximately equal to the ratio of the lengths.

The excitation threshold of the trigger is so chosen that the trigger will fire only for pulses of length τ_l .

The remaining assemblies of the receiver are duplicates of units in the transmitter and, hence, will not be described again.

For decoding, the distributor D is controlled by a pulse generator which functions synchronously with the

generator on the transmitting end. Synchronization is implemented by the synchronizing pulses (markers) and by subsidiary pulses entering into the code combinations.

Moreover, start-stop distributor-synchronizing pulses, obtained by use of the "void" distributor elements I-VI and the preparatory first element B, which discriminates the synchronizing pulse from the code pulses, are provided on the receiving end.

Code and synchronizing pulses entering the channel amplified and clipped, impinge on one of the windings of magnetic memory element K and prime it every time. Sensing of element K is executed under control of the multivibrator simultaneously with the switching of the magnetic distributor 180° out of phase with the code pulses. Thus, after a synchronizing pulse impinges on the load winding of element K, a pulse is formed, simultaneously with the formation of a pulse at element B. If the first code pulse to arrive were 2^5, then in the next half period, pulses would be formed both at the load winding of K and at the winding of element 1, etc. The interaction of these elements is shown in Fig. 10. Winding W_{L2} of element K and all distributor elements I through VI would be switched in the reverse direction.

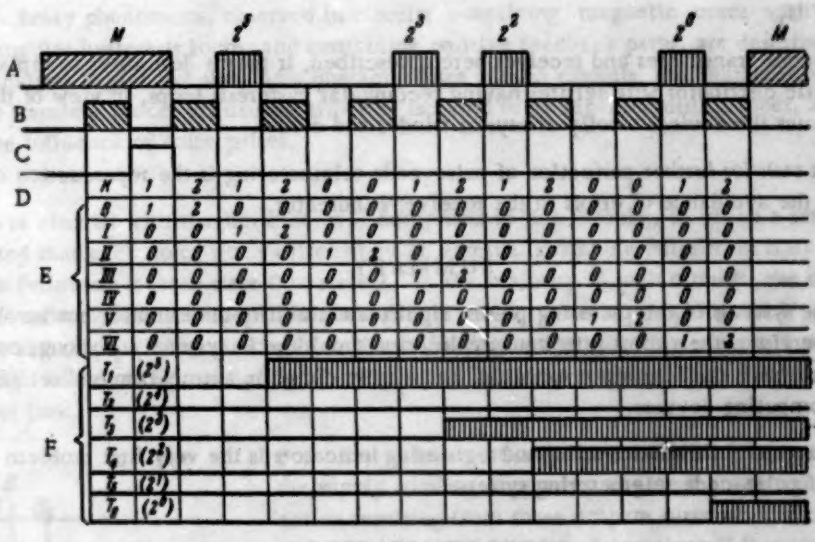


Fig. 11. Cycle diagram for receiver functioning.
A) Code pulses; B) Control Pulses; C) Selector pulses; D) Switching of the memory elements; E) Rectifier switching; F) Trigger switching.

Winding W_{L1} is an element of coincidence block, CB. If a memory element were primed by a code pulse, then sensing would cause a pulse to be formed only at distributor element winding W_{L1}, but would suppress pulse formation at winding W_{L2}. Thus, a pulse would be present at input 2 of the corresponding trigger. This trigger, in switching, would unblock a definite tube in the conductor assembly and, in the circuit of the indicating device, a current would be established corresponding to the received code pulse. The circuit would function similarly to the end of the cycle, at which time the indicating device would have a current corresponding to the received code. If there were no variation in the measured quantity, then no switching would occur in the circuit and the magnitude of the current in the indicator would not change. For a new combination of code pulses, a series of triggers would, in general, be switched; triggers not taking part in this decoding would be restored to their original state by pulses formed at winding W_{L2}. Here, pulse suppression would not occur, since, in this case, the elements of K are not switched. No pulse will appear at input 2 since there will be no pulse from winding W_{L1} of element K to enable sensing in the coincidence circuit. It should be explained that if, in the receiving process, the code improperly impinges on the distributor, due either to a transient disturbance in the circuit or to an inacceptably large magnitude of instantaneous noise, the indicating device will not be affected, since the cycle repeats itself after 0.2 sec and the actual value of the measured quantity will be restored.

The functioning of the receiver is shown in Fig. 11.

The overall error of the design is comprised of the discreteness error, $\delta_D \leq \pm[2(2^n-1)]^{-1}$ (where n is the

number of elements in the code) and the summing error, δ_s, affected by the variation in voltage drop across the conductor circuit resistors, which in turn depends on the combination of branches switched. Due to this, the ratio of the k'th current i_{2^k} to the current i_{2^0} will deviate from the law of geometric progression. The ratio i_{2^k}/i_{2^0} is supposed to equal 2^k, i.e., $i_{2^1}/i_{2^0} = 2$, $i_{2^2}/i_{2^0} = 4$, ... $i_{2^{n-1}}/i_{2^0} = 2^{n-1}$. The deviation of current magnitude from the values defined by this law, relative to the quantity $2^n - 1$, will give rise to an error $\delta_{s(k)} = \left(\frac{i_{2^k}}{i_{2^0}} - 2^k\right)/(2^n - 1)$, where $\underline{k}$ is the ordinal number of the term in the binary series.

The currents, i_{2^k} and i_{2^0}, are expressed by elementary formulas and their computation presents no complexity.

The maximum summing error for the design will occur when all branches of the conductors are switched: $\delta_s = \sum_{1}^{n} \delta_{s(k)}$. In the design considered here, this error amounts to 0.55 %.

The influence of a series of other factors on the accuracy of measurement, e.g., the inaccuracy of calibrating the resistors in the summing conductor circuit, the instability of the voltage source for the conductor circuit, etc., may be similarly estimated and held to a minimum. We might mention that it is possible to ignore variation in the voltage supplying the equalization circuit if its source is the same as the source of the voltage used in the rectification of U_1.

In the pulse-code transmitter and receiver herein described, it seems desirable to replace the ribbon material cores in the magnetic distributor with ferrites having rectangular hysteresis loops, in view of their improved technology, or to construct the devices wholly of crystal triodes and diodes.

An important task for further perfection of pulse-code telemetering is the reproduction of digital indications, which would allow the avoidance of errors in the receiver's indicator.

SUMMARY

1. Pulse-code systems of telemetering possess significant stability under noise, but involve relatively complex apparatus. Therefore, their most felicitous application should be in systems with long, complex links, where the use of other systems is not feasible or possible, and also in complex automation and remote control units in conjunction with computing devices.

2. The construction of digital reading and registering indicators is the very first problem to be solved in the further perfection of pulse-code telemetering systems.

LITERATURE CITED

[1] "Oxford pulse-code modulation system," Proc. Inst. Elec. Engrs, V. 41, No. 7 (1953).

[2] "Coding by feedback methods." Proc. Inst. Elec. Engrs. V. 41, No. 8, (1953).

[3] Electrical Communications, V. 24, No. 3, (1947).

[4] Principles of Implementing Links and Remote Control in EVS* Systems [In Russian]. Appendix 4. Pulse-code telemetering systems. Report of Central Scientific Research Electrotechnical Laboratory, Ministry of Electric Power Plants (USSR), 1954.

[5] V. A. Zhozhikashvili and K. G. Mitiushkin, "On the functioning of computing-switching designs containing elements with rectangular loops" Automation and Remote Control (USSR), 16, 4 (1955).

Received May 31, 1957

* Transliteration of Russian — Publisher's note.

RELAY PHENOMENA IN LOOP CIRCUITS CONTAINING MAGNETIC CORES WITH RECTANGULAR HYSTERESIS LOOPS

V. A. Zhozhikashvili and K. G. Mitiushkin

(Moscow)

Relay phenomena, observed in circuits containing magnetic cores with rectangular hysteresis loops, and containing positive feedback paths, are described. An analysis is given of the static characteristics of such circuits, the character of their transient processes, caused either by isolated or by repeated disturbances, and of the influence of noise pulses.

A large group of circuits with magnetic cores is comprised of link circuits, in which a series of identical cores are so connected that each core, when switched by an external switching pulse from state 1 to state 0 (Fig. 1), switches the core following it from state 0 to state 1. In the majority of such circuits, the cores are connected according to one or another of the schemes shown in Fig. 2, a and b. Windings W_1, W_2 and W_D are called, respectively, output, input and supply (shift) windings. A circuit joining two neighboring cores and consisting of windings W_1 and W_2, resistor r_u and valves B_1 and B_2 (Fig. 2, a) or valve B_1 and capacitor C (Fig. 2, b) is called a connecting link, or a transfer link.

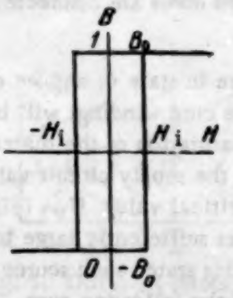


Fig. 1. Rectangular hysteresis loop.

Let us assume that an external switching pulse, flowing through the supply winding, engenders negative ampere turns in the core. The action resulting from these ampere turns we will call voiding of the core. Voiding will be termed complete if it proceeds along the descending branch of the hysteresis cycle and, during the time of action of the switching pulse, the core is switched from state 1 to state 0.

In contradistinction to an external switching pulse, a switching pulse entering from the connecting link to the input core winding engenders in it positive ampere turns. The switching thereby induced we call priming of the core. We shall say that the priming is complete if, during the time of action of the switching pulse, the core state changes from point 0 to point 1 along the ascending arm of the complete hysteresis loop.

Voiding of a core is accompanied by the appearance of a pulse in its output winding, the power of which is the larger, all other conditions being equal, the larger the power in the supply winding pulse. For a sufficiently large power in the supply winding pulse, the power in the output winding pulse, during voiding of the core, may be significantly greater than the power necessary to prime the following core. In such a regimen, each core is a pulse amplifier. If the losses in the connecting link are not too large, then the complete voiding of each core is accompanied by the complete priming of the following core. Thus, the process of voiding and priming is transferred from core to core, and bears this sustained character for any number of cores in the chain. This process is terminated with the voiding of the final core in the chain.

If a connecting link is furnished between the last and the first cores comprising the circuit, we shall then have a series of connected pulse amplifiers with a positive feedback connection. In the following, we shall call such a circuit a looped circuit.

A looped circuit with magnetic cores may, if the supply windings are fed by a source of switching pulses, possess many stable states. A transition of the circuit from one stable state to another occurs at definite critical values of voltages, currents and circuit parameters, and has a snowballing character.

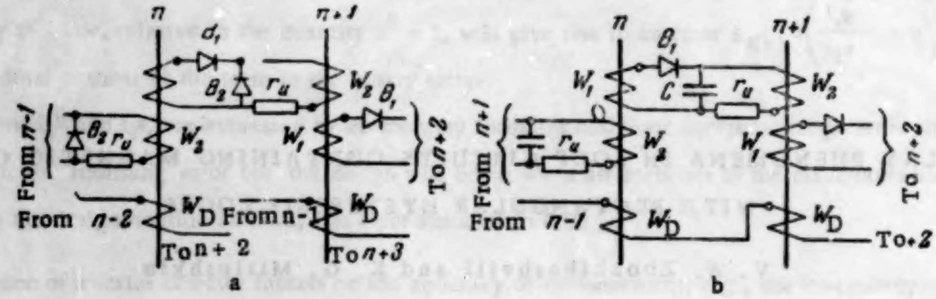


Fig. 2. Two types of interconnections.

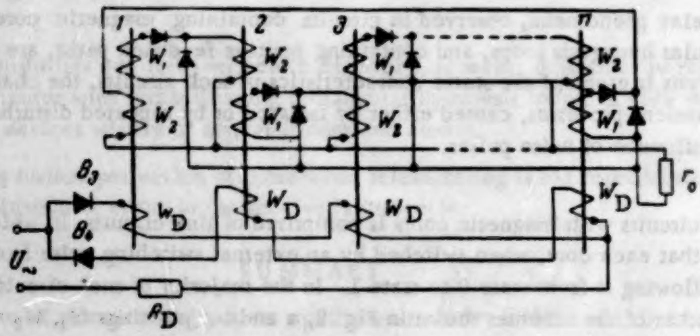


Fig. 3. A two-half-period looped circuit.

We consider the looped circuit depicted in Fig. 3. This is a development of the circuit shown in Fig. 2, a, and henceforth will be referred to as two-half-period. In this circuit, the supply windings of the odd-numbered cores are connected in one series, and the supply windings of the even-numbered cores are connected in another series.

We assume that initially the supply voltage $U_{\sim} = 0$ and that all cores are in state 0, and we call this the first stable state. If $U_{\sim}$ is now slowly increased, then at the terminals of all the core windings will be observed a gradual increase in the amplitude of noise pulses, stemming from non-ideal saturation of the material in the neighborhood of point 0, and also from the presence of reverse conductance in the supply circuit valves, B_3 and B_4, which leads to partial priming of the cores. When voltage $U_{\sim}$ attains the critical value U_{V1} (Fig. 4) which may be called the excitation voltage, the amplitude of the noise pulses becomes sufficiently large to trigger the snowballing transfer of the circuit from the first stable state to the second. In this state each source pulse voids one of the cores, while complete voiding of a core entails complete priming of the following core. Thus, in the second state, each of the cores voids once per cycle, a cycle consisting of a number of half-periods in the supply circuit equal to the number of cores in the chain. A transition of a circuit from the first to the second stable state without special priming pulses will be called a self-excited circuit. In self-excited circuits, generally speaking, the amplitude of the noise pulses is attenuated, since a significant portion of the supply voltage is applied in voiding the cores. This allows the supply voltage to be significantly increased without disrupting the functioning of the circuit. When the supply voltage is made equal to U_{V2} the amplitude of the noise pulses again attains the earlier value and, as a result, a second snowballing process occurs, implying a transition of the circuit to the third stable state, characterized by the fact that with each pulse in the supply circuit two cores are simultaneously voided. With additional increases in voltage, further skips may be obtained.

In contradistinction to the transition to the second stable state, the transition of the circuit to the third, fourth and succeeding states is brought about not only by noise pulses but also by a reverse influence of a voiding core on the previous core.

After the circuit has been transferred to the second stable state by voltage U_{V1}, it can remain in that state

even if the supply voltage is significantly reduced. Fig. 4 shows that only at voltage U_{rel} will the circuit return to its original, nonexcited state. There is thus the typical loop characteristic, peculiar to relay devices.

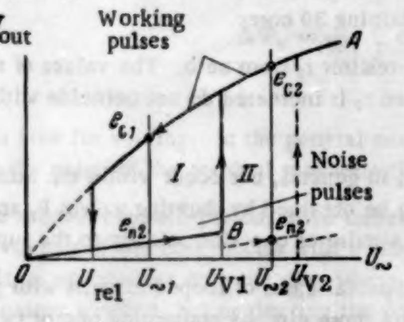


Fig. 4. Looped circuit characteristics. OA is the operating pulse curve; OB is the noise pulse curve; U_{V1} and U_{V2} are supply circuit voltages, corresponding to transitions to the first and second excited states; U_{rel} is the relaxing voltage, corresponding to a circuit transition from an excited to a non-excited state; e_{c1} and e_{c2} are the currents corresponding to the operating pulse amplitudes in the operating regions I and II; e_{n1} and e_{n2} are the currents corresponding to the noise amplitudes in the same cases; U_1 and U_2 are values of supply voltage.

The greatest practical interest is presented by the characteristics of regions I and II. If the supply voltage is chosen to lie within the limits of region I($U_{rel} < U_\sim < U_{V1}$), then a self-excited looped circuit is impossible. In order that the circuit be transferred to an excited state, pulses from an external source are necessary to prime one of the cores. Priming may be complete or incomplete. In the latter case, however, the variation in flux during priming must be larger than a certain minimum quantity necessary for exciting the circuit. As will be made clear in the sequel, the length of the transient processes will depend on the magnitude of the flux variation during priming.

If $U_\sim = U_{\sim 1}$, then the amplitude of the noise pulse for any core is equal to e_{n1}. After priming one core, the variable signal amplitude for core voiding equals e_{c1}.

If the supply voltage is chosen within the limits of region II ($U_{V1} < U_\sim < U_{V2}$), then the introduction of supply voltage will produce a self-excited circuit. This will automatically establish a regimen in which one of the cores is found in state 1, and this state will be transferred from core to core. All other cores will remain in state 0. If, for example, $U_\sim = U_{\sim 2}$, then the available signal amplitude will equal e_{c2}, and the amplitude of the noise pulses will equal e_{n2}.

Both modes of functioning of the looped circuit, i.e., with self-excitation (region I) and without self-excitation (region II), may have interesting practical applications.

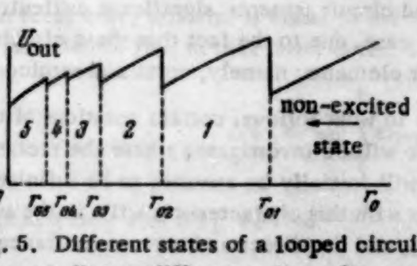


Fig. 5. Different states of a looped circuit, corresponding to different values of r_0.

Curves OB and OA, bounding the characteristic loop respectively below and above, may be obtained in the following way. Let one core be wound according to the data calculated for the cores in the looped circuit, and let its windings, W_1 and W_2, be loaded as they would have to be loaded in the circuit, and then include winding W_D together with the valves and resistor R_D, after which increasing the supply voltage for this single core will give curve OB. If the same procedure is repeated, but with shorted supply circuit valves, then curve OA will be obtained. A somewhat higher free-running characteristic may be obtained if the circuit for windings W_1 and W_2 is open.

Families of loop characteristics, analogous to those shown in Fig. 4, and derived for different circuit parameters, core windings, etc., give a sufficiently clear picture of the static properties and functioning of looped circuits. These characteristics are a reflection of the excitation voltage, the releasing voltage, the source pulse amplitudes and the noise pulse amplitudes.

Different stable states may be obtained, not only by varying the magnitude of the supply voltage, but also by different values of the circuit parameters. These states differ one from another by the number of voided cores during the time of one cycle. Thus, for a very large resistor r_0 in the connecting link, none of the cores is primed, and the circuit will be in the non-excited state (Fig. 5). With the lowering of resistance r_0 to a value r_{01}, one of the cores will be primed, as a result of amplification of noise pulses. The circuit will transfer to the first excited state, in which each of the cores will be voided once per cycle. This regimen is maintained until r_0 is lowered

to a value r_{02}. At this latter value, the circuit transfers to the second excited state, in which each of the cores is voided twice per cycle. With further decrease of r_0, one may obtain the third, fourth and all succeeding excited states, up to the final state, in which each core is voided as many times as the period of source frequency can be packed into one working cycle of the circuit. In the case shown in Fig. 5, the circuit contained 10 cores. The existence of all stable states has been observed in a circuit containing 30 cores.

All states of a circuit may also be obtained by increasing resistor r_0 from zero. The values of resistor r_0 at which the circuit transfers from one excited state to another when r_0 is increased do not coincide with the values of r_0 when it is decreased.

In the case of cores with little noise, self-excitation may, in general, not occur within the attainable limits of supply voltage and current. In such cases, self-exitation may be obtained by shunting valves B_3 and B_4 with a large resistance, thereby causing a partial priming of a core by a pulse of opposite polarity to the supply pulses.

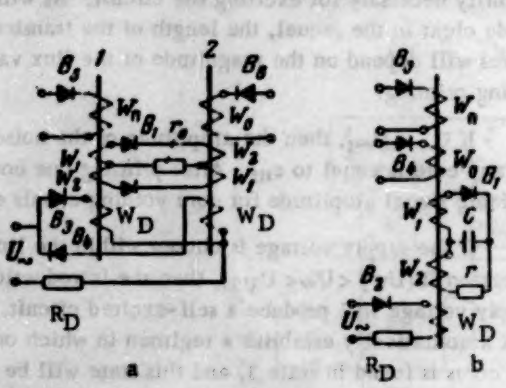


Fig. 6. Two dynamic trigger circuits.

Special cases of looped circuits with positive feedback are those circuits containing one or two cores. Such circuits may be utilized as magnetic dynamic triggers or relays. Two such circuits are shown in Fig. 6, a and b. These circuits possess two stable states, and the loop characteristics, analogous to those given in Fig. 4, may be obtained with them. In the circuit with one core, the voiding of the core leads to the charging of the capacitor. Priming of the core occurs as a result of the condenser discharging through input winding W_2. In the circuit with two cores, priming of one core occurs as a consequence of voiding of the other.

To function as a control relay, the core is furnished with a starting winding W_s and an inhibiting winding W_i. The approximate characteristics of a control relay for two cores are given in Fig. 7. In comparison with other circuits, the magnetic non-contact relay circuit given here possesses both advantages and disadvantages, but an investigation of these questions is not within the scope of the present paper.

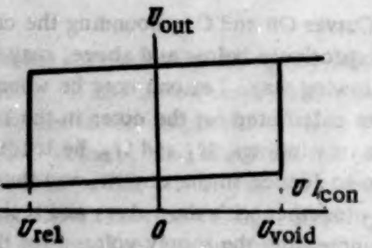


Fig. 7. Characteristics of a control relay.
U and I_{con} are the voltage and current in the control path; U_{void} and U_{rel} are the voltages in the control path corresponding to the voiding and releasing relays; U_{out} is the voltage of the relay output.

An investigation of the transient processes in a looped circuit presents significant difficulty in the general case, due to the fact that these circuits contain nonlinear elements, namely, cores and semiconductors.

In what follows, certain questions of transient response will be investigated where the rectangular hysteresis loop will initially be assumed to be infinitely narrow (cores with this characteristic will, in the sequel, be called ideal), and the direct and reverse resistance of the valves will be considered, respectively, as being constant and infinite.

In view of the fact that in the steady state each core of the looped circuit must be switched around the complete hysteresis loop in the course of one half-period of supplying voltage, both for voiding and for priming, and investigation of the transient processes in a looped circuit is most felicitously carried out separately for each supply half-period. Such an approach was recently employed in investigating the workings of magnetic amplifiers [1].

We shall consider looped circuits with arbitrary numbers of cores. For definiteness, we will take them in the form of two-half-period circuits.

At the input winding W_2 of some core, in one of the half-periods when there is no current in the supply wind-

ing of this core we supply a voltage pulse, $e_C = E_C \sin \omega t$, of duration t_C where t_C may have any value from $t_C = 0$ to $t_C = T/2$ (T being the period of the supply voltage). This pulse gives rise to a variation $\Delta\Phi_C$ of the core flux equal to

$$\Delta\Phi_c = \frac{1}{W_2}\int_0^{t_c} e_c\,dt = \frac{1}{W_2}\frac{E_c}{\omega}(1-\cos\omega t_c),$$

priming this core for voiding. In the general case, priming may be either incomplete or complete and $\Delta\Phi \leq 2\Phi_0$ where Φ_0 is the value of the residual magnetic flux in the core.

In the succeeding half-period, when current begins to flow in the supply winding W_D of the core under consideration, voiding of the core occurs with a change of flux $\Delta\Phi_1$ during voiding which is generally less than $\Delta\Phi_C$. In what follows we consider only the case where $\Delta\Phi_1 = \Delta\Phi_C$. This will always be the case if the amplitude of the emf in the source winding, but neglecting the voltage drop across resistor R_D, and the amplitude of the source voltage satisfy the condition

$$\frac{E_\sim}{W_D\omega} \geqslant 2\Phi_0. \tag{1}$$

The time t_1 necessary for voiding the core is determined from the following equalities:

$$\frac{1}{W_D}\int_0^{t_1} e_\sim\,dt = \frac{E_\sim}{W_D\omega}(1-\cos\omega t_1) = \Delta\Phi_1. \tag{2}$$

During the time of voiding, a pulse is induced in the output winding W_1 with emf

$$e_{out} = E_{out}\sin\omega t = \frac{W_1}{W_D}E_\sim \sin\omega t,$$

and length t_1.

For an ideal core, this emf is equal to the voltage approached in the winding W_2 of the following core, inducing a variation of flux in that latter core. This flux variation $\Delta\Phi_2$ equals:

$$\Delta\Phi_2 = \frac{1}{W_2}\int_0^{t_1} e_{out}\,dt = \frac{W_1E_\sim}{W_2W_D\omega}(1-\cos\omega t_1).$$

Comparing $\Delta\Phi_2$ with $\Delta\Phi_1$, we find

$$\Delta\Phi_2 = \Delta\Phi_1\frac{W_1}{W_2} = \Delta\Phi_1 k_W. \tag{3}$$

Thus, k_W shows itself to be the transfer coefficient (amplification factor) of the flux and, depending on whether this coefficient is greater than, or less than, unity, there will be, respectively, an increase or a decrease in the magnitude of flux variation and in the voiding time for each successive half-period. For $k_W > 1$, the process grows up to the point where the voiding of the core in some half-period produces full priming of the following core (Fig. 8).

With this latter half-period begins the steady state, during each half-period of which each core in turn is voided and the core following is completely primed. For $k_W < 1$, the process, conversely, decays until the flux variation during priming becomes equal to zero.

If we number the cores and the supply half-periods, counting as number one that half-period in which the signal was applied to the first core, then the flux variaton in the second core during its second-half-period priming will equal $\Delta\Phi_2 = k_W\Delta\Phi_1$; the flux variation in the third core during the third half-period will equal

$\Delta\Phi_3 = k_W \Delta\Phi_2 = k_W^2\, \Delta\Phi_1$, etc. In general, the flux variation will equal $\Delta\Phi_n = k_W^{n-1}\, \Delta\Phi_c$, where $n \geq 1$. The core voiding time can be determined in the following fashion:

$$\Delta\Phi_n = \frac{1}{W_D} \int_0^{t_n} e_\sim\, dt = \frac{E_\sim}{W_D\,\omega} (1 - \cos\omega t_n) = k_W^{n-1}\, \Delta\Phi_c, \tag{4}$$

from which

$$\cos\omega t_n = 1 - \frac{k_W^{n-1}\, \Delta\Phi_c\, W_D\, \omega}{E_\sim}. \tag{5}$$

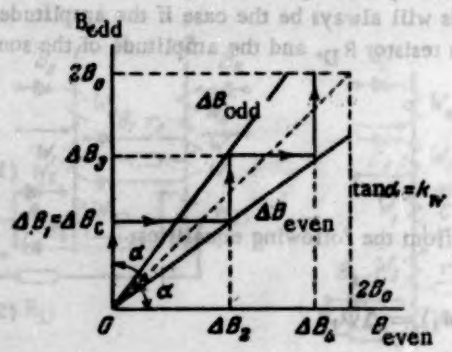


Fig. 8. Process of exciting a looped circuit when $k_W > 1$. ΔB_c is the flux variation in the first core, induced by signal pulses; ΔB_1 and ΔB_3 are the flux variations in the first and third cores for priming, ΔB_2 and ΔB_4 the same quantities for the second and fourth cores; ΔB_{odd} and ΔB_{even} are the flux variations in the odd and even cores.

In order to determine the number of half-periods in the transient process, and the number of the core with which complete voiding begins (steady-state), it is necessary to make the substitution $\Delta\Phi_n = 2\Phi_0$ in Equation (4). We then obtain

$$k_W^{n-1}\, \Delta\Phi_c = 2\Phi_0,$$

from which

$$n_{ss} = \frac{1}{\log k_W} \log \frac{2\Phi_0}{\Delta\Phi_c} + 1. \tag{6}$$

Correspondingly, the voiding time in the steady state is obtained from Expression (5):

$$\cos\omega t_{ss} = 1 - \frac{2\Phi_0\, W_D\, \omega}{E_\sim}. \tag{7}$$

Thus, as follows from (3), the phenomenon of self-excitation is possible only for $k_W > 1$, which is the necessary condition for the existence of relay-like jumps and stable functioning in several different states of the circuit we have been considering. It must hold in all those cases when cores are utilized in commutator, counter, accumulator and relay circuits.

For $k_W < 1$, a self-exciting looped circuit is impossible but, due to the positive feedback, it is possible to evoke a significant amplification of noise pulses. In fact, if we assume that in the circuit of Fig. 6, a, during the half-period when current is flowing in winding W_D of core II but current is blocked from winding W_D of core I by valve B_3, a noise pulse might act in winding W_n a pulse given by $e_n = E_n \sin\omega t$, where t may take any value in the range $0 < t < T/2$. The first such pulse would occasion a partial priming of core I, varying the flux in the core by the quantity $\Delta\Phi_n$. In the following half-period, core I would be voided and core II would be primed, the flux variation in the latter core being $k_W \Delta\Phi_n$, in accordance with the above. The voiding of core II leads in turn to a flux variation in core I of magnitude $k_W^2\, \Delta\Phi_n$. Adding in the action of the noise, we obtain a total flux variation in core I of $\Delta\Phi_n + k_W^2 \Delta\Phi_n$. In the succeeding voiding of core I, a flux change of $k_W(\Delta\Phi_n + k_W^2\, \Delta\Phi_n)$ is induced, etc.

Continuing this line of reasoning, we do not find it difficult to establish that, in the n'th half-period after the appearance of noise pulse e_n, the flux change induced in core II will be

$$\Delta\Phi_{II} = \Delta\Phi_n (k_W + k_W^3 + k_W^5 + k_W^7 + \ldots + k_W^n), \tag{8}$$

where n is an arbitrary odd number.

If $k_W < 1$ then, for $n = \infty$, we obtain the steady-state value of the flux in core II:

$$\Delta\Phi_{IIss} = \Delta\Phi_n \frac{k_W}{1 - k_W^2}. \qquad (9)$$

The flux change in core II at the $(n + 1)$th half-period will be equal to

$$\Delta\Phi_I = \Delta\Phi_n (1 + k_W^2 + + k_W^4 + k_W^6 + \ldots + k_W^{n+1}). \qquad (10)$$

The steady-state flux in core I equals

$$\Delta\Phi_{Iss} = \Delta\Phi_n \frac{1}{1 - k_W^2}. \qquad (11)$$

Thus, the steady-state flux defining the resulting flux change is proportional to the flux induced by the noise pulse but, in the first core, is larger than this initial flux by the factor $\frac{1}{(1 - k_W^2)}$. This leads to an amplification of the noise pulses taken off the output winding of core I, as compared to the same pulses in the absence of positive feedback. Fig. 9 shows the variation of magnetic induction in cores I and II for $k_W < 1$, corresponding to the process just described.

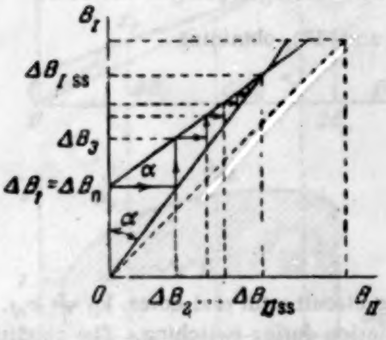


Fig. 9. The process of trigger excitation under the influence of noise pulses for $k_W < 1$. B_I and B_{II} are, respectively, the inductions in cores I and II, and ΔB_n are the variations of induction with each noise pulse.

In circuits consisting of real magnetic cores, having a coercive force $H_l \neq 0$, the transition from the non-excited state to the excited state may occur only if the signal pulse e_c or the noise pulse e_n exceeds a definite threshold value.

In order to determine the aforementioned threshold value and to elucidate the conditions under which it is possible for the circuit to function in excited states with completely voided cores, we shall consider not infinitely narrow hysteresis loops, but rectangular loops with a finite width $2H_l$. Here we will take the circuit to be of the form given in Fig. 3.

We introduce the following notation: $a = \left(\frac{W_D^2}{R_D} + \frac{W_2^2}{r_0}\right)$, $b = S \cdot 10^{-8}$, where S is the cross-sectional area of the core, $c = \frac{L}{0.4\pi}$, where L is the average length of magnetic line,

$$A = \frac{U_{\sim} W_D}{R_D ab},$$

$$D_1 = \frac{cH_c}{ab}(1 + k_W), \quad D_2 = \frac{k_W cH_c}{ab}(1 + k_W) + D_3, \quad D_3 = \frac{cgH_c}{n},$$

$$g = \frac{r_0}{W_2^2}, \quad \frac{\Delta B_2}{\Delta B_1} = k_B,$$

where ΔB_1 and ΔB_2 are the total variation in inductance during the voiding and the priming of the core, respectively.

For a sinusoidal source of pulses, the process of switching cores may be thus described:

$$\frac{dB}{dt} = A \sin \omega t - D_1.$$

The process of varying the induction during priming, under the same conditions, may be written as:

$$\frac{DB}{dt} = k_W A \sin \omega t - D_2.$$

Integrating, we get

$$\frac{A}{\omega}(1-\cos\omega t_1)-D_1t_1=\Delta B_1, \tag{12}$$

$$\frac{k_W A}{\omega}(1-\cos\omega t_1)-D_2t_1=\Delta B_2, \tag{13}$$

where t_1 is the time during which the induction varies by the amount ΔB_1 during voiding of the core.

As in the case of the ideal cores, we compare Expressions (12) and (13), obtaining

$$\Delta B_2=k_W\,\Delta B_1-D_3t_1$$

or

$$k_B=k_W-D_3\frac{t_1}{\Delta B_1}. \tag{14}$$

Thus, in contradistinction to the circuit with ideal cores, in the circuit with real cores, $k_B \neq k_W$. The value of k_B depends on the ratio of the switching time to the induction variation during switching. The condition $k_W > 1$ is no longer sufficient to guarantee the possibility of jumps, and of the circuit functioning in excited states.

Reasoning the same way as for the case of ideal cores, it is possible to state that if in one of the half-periods when current is absent from the source winding of the first core, priming of this core is occasioned by a pulse e_C of magnitude ΔB_C, then in the following half-period, when the first core voids, the flux variation ΔB_1 will in general be less than or equal to ΔB_C. As in the ideal case, we will consider only the situation when $\Delta B_1 = \Delta B_C$. For this, the following condition must hold:

$$\frac{2A}{\omega}-D_1\frac{T}{2}\geqslant 2B_0.$$

We assume that the time of voiding t_1 is determined by Expression (12). Let $t_1 = f_1(\Delta B_1)$. Voiding of the first core entails priming of the second core, wherein the induction variation is determined from (13):

$$\Delta B_2=k_W\,\Delta B_1-D_3t_1=k_W\Delta B_1-D_3f_1(\Delta B_1).$$

In the second core, during the following half-period, voiding time is determined from the equation

$$\frac{A}{\omega}(1-\cos\omega t_2)-D_1t_2=\Delta B_2,$$

from which

$$t_2=f_1(\Delta B_2)=f_1[k_W\,\Delta B_1-D_3f_1(\Delta B_1)]=f_2(\Delta B_1).$$

When the second core voids, the induction variation in the third core equals

$$\Delta B_3=k_W\,\Delta B_2-D_3t_2=k_W^2\,\Delta B_1-k_W\,D_3\,f_1(\Delta B_1)-D_3f_2(\Delta B_1).$$

Reasoning similarly, we may write for the voiding of the n'th core during the n'th half-period:

$$\Delta B_n=\frac{A}{\omega}[1-\cos\omega f_n(\Delta B_1)]-D_1f_n(\Delta B_1).$$

The induction variation in the (n + 1)th core during the n'th half-period may be determined from the expression:

$$\Delta B_{n+1}=k_W^n\,\Delta B_1-k_W^{n-1}\,D_3f_1(\Delta B_1)-k_W^{n-2}\,D_3f_2(\Delta B_1)-\ldots$$
$$\ldots-k_W\,D_3f_{n-1}(\Delta B_1)-D_3f_n(\Delta B_1).$$

Thus, although the analytic investigation of the transient processes is theoretically possible, in practice, it involves considerable difficulty.

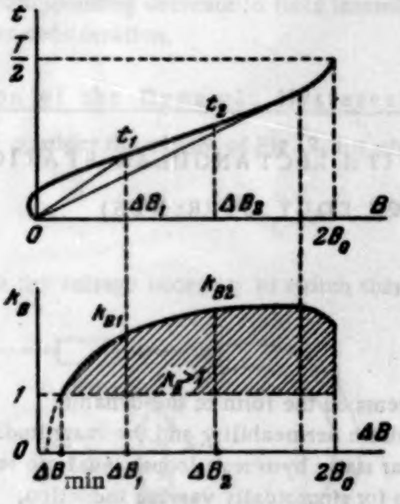


Fig. 10. Consturction of the curve, $k_B = f(\Delta B)$.

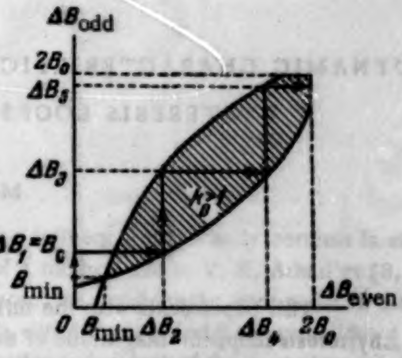


Fig. 11. Excitation process of a looped circuit for sinusoidal switching pulses, and $H_l \neq 0$.

A graphic determination of the conditions for exciting the circuit, and of the length of the transient processes, may be carried out as follows. From Equation (12), the dependence of $\underline{t}$ on ΔB is constructed, as shown in Fig. 10; above this curve, using it and Equation (14), one constructs the dependence of k_B and ΔB. Functioning in an excited state is possible only on the condition that, for values of ΔB lying in the interval $B_{min} < \Delta B < 2B_0$, k_B assumes values greater than unity. The value of B_{min} for which $k_B = 1$ thus defines the minimum signal necessary for excitation of the circuit. If the magnitude of the signal is known, then the length of the transient process may be determined from the curve of k_B. We do this by determining ΔB_1 from the known signal amplitude. Using this value of ΔB_1, we obtain k_{B_1} from the curve. Multiplying ΔB_1 by k_{B_1}, we obtain a new value ΔB_2 and so on, until the point is reached at which $\Delta B \geq 2B_0$.

In Fig. 11, the region of relay action is shaded. On the axes of the abscissas and the ordinates lie the induction values for, respectively, even and odd numbers of cores.

SUMMARY

1. Looped circuits with magnetic cores possess, generally speaking, many stable states wherein the transition of the circuit from one stable state to another occurs for definite values of current and voltage, and has a snowballing character. This phenomena stems from the amplifying action of the cores and from the presence of positive feedback.

2. The stable states, corresponding to regions I and II of Fig. 4, may be used to obtain counting-switching devices, working under a regimen either of external excitation (region I) or of self-excitation (region II).

3. A looped circuit consisting of ideal cores ($H_l = 0$), with the condition that $k_W > 1$, may not, in principle, remain in a non-excited state, since an arbitrarily small disturbance suffices to transfer the circuit to an excited state.

4. A looped circuit consisting of real cores ($H_l \neq 0$) may be taken out of the non-excited state providing that the disturbing action of signal or noise produces a switching of a core which exceeds a certain threshold value, ΔB_{min}.

LITERATURE CITED

[1] R. A. Ramey, "On the Mechanics of Magnetic Amplifier Operation," Trans. Am. Inst. Elect. Engrs., V. 70, Part 2, (1951).

Received March 11, 1957

DYNAMIC CHARACTERISTICS OF CORES WITH RECTANGULAR STATIC HYSTERESIS LOOPS (INFLUENCE OF EDDY CURRENTS)

M. A. Rozenblat

(Moscow)

This paper deals with the influence of eddy currents on the form of the dynamic hysteresis loop, the magnitude of the differential magnetic permeability and the magnitude of the dynamic coercive force of cores with rectangular static hysteresis loops. Analytic representations are found for the dynamic hysteresis loop for sinusoidally varying induction, for sinusoidally varying field intensity, and for dc magnetization of the core.

Experimental data are cited in confirmation of the theoretical results.

INTRODUCTION

For the cores of magnetic amplifiers, widespread use is made of alloys with rectangular hysteresis loops (e.g. 50NP, 65NP, N34K29MZP, etc.), which makes possible significant improvement in linearity, increase in gain, and decrease in lag of amplifiers with positive feedback [1]. Such alloys have also found wide application for other electromagnetic devices, for example, in magnetostatic triggers, in memory elements, and in pulse transformers, etc.

In the function of such devices, the width and form of the dynamic hysteresis loop are significant. For example, the magnitude of the differential magnetic permeability, $\mu_D = dB/dH$, on the "vertical" portions of the dynamic hysteresis loop determines the value of the gain for a number of types of magnetic amplifiers with positive feedback.

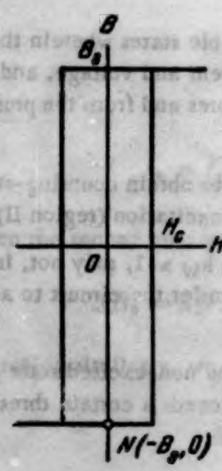


Fig. 1

In order to appraise the quality of cores with magnetic hysteresis loops, one frequently checks their dynamic characteristics experimentally (e.g., by means of an oscillograph). To interpret the measurement results, it is necessary to know the dependence of the dynamic characteristics on the conditions of measurement.

The aim of this paper is the theoretical investigation of the influence of eddy currents on the dynamic characteristics of toroidal cores with rectangular static hysteresis loops under different conditions of magnetization. We ignore the influence of magnetic viscosity. For strip material about 0.03 to 0.04 mm and above in thickness, this influence is ordinarily small in comparison with the effect of eddy currents [2]. It is assumed that in the limit the static hysteresis loop has the ideal rectangular shape (Fig. 1), characterized by the residual inductance B_r being equal to the saturation induction B_s and the differential permeability μ_D being equal to infinity for values of $|B| < B_s$.

We might mention that, for a number of cores of alloys 65NP and N34K29MZP, the ratio B_r/B_s attains a value of 0.95-0.99, and μ_D a value of 10^7 gauss/oersted.

In a dynamic regime, the cores and their lamina or turns are magnetized inhomogeneously across their cross-section. Therefore, we shall understand by the dynamic characteristics of a core the average dependence, across

its cross section, of the magnetic induction on the strength of the field induced by the current in the magnetizing winding. In the present work, we assume that the ratio of the inner and outer core diameters is close to unity, and we therefore ignore the inhomogeneity of the core magnetization, due to the lengthening of the magnetic path, and the corresponding decrease in field intensity, stemming from the increase of diameter in the portion of the core under consideration.

Equation of the Dynamic Hysteresis Loop.

We consider the circuit of Fig. 2, for which we have

$$w\frac{d\Phi}{dt}10^{-8}+Ri=e, \tag{1}$$

where e is the voltage necessary to switch the core.

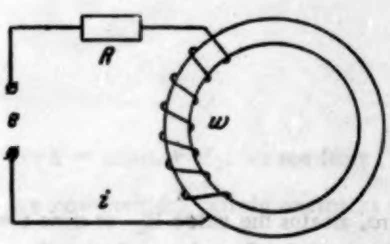


Fig. 2

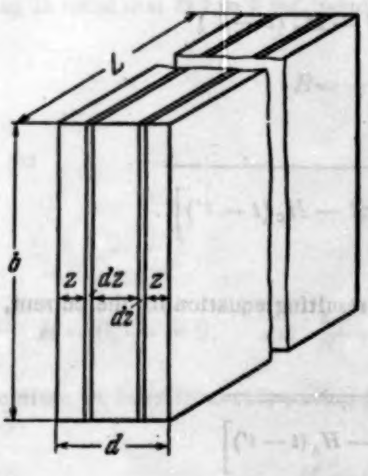


Fig. 3

The influence of the eddy currents is studied by means of a method due to V. K. Arkad'ev [3, 4]. For this purpose, we consider the core to be in the form of a lamina with length L, height b and width d (Fig. 3). For the strips (windings) of the core, $L = nl$, where n is the number of turns about the core, l is the length of one turn, b is the core depth and d is the width of the strip; as a rule, $L \gg b \gg d$.

We suppose that the initial magnetic state of the core corresponds to the point N $(-B_s, 0)$ of the hysteresis loop (Fig. 1). If the field intensity, induced in the core winding w by the current i, exceeds the coercive force H_c, then the induction in an infinitesimal strip of width dz is changed instantaneously from $-B_s$ to $+B_s$. The flux change in the lamina begins at the circumference and proceeds inward.

Let the flux penetrate to depth z. Then, further variation of the induction in the two strips of width z (Fig. 3) induces an emf (taking $d \ll b$), $e = -2B_s \cdot 2b\frac{dz}{dt}10^{-8}$. In a strip of width z, this emf induces an eddy current, $i_f = eY$, where Y is the conductivity of this strip, the magnitude of which is found from the formula $Y = \frac{z\gamma L}{2b}$, γ being the specific conductivity.

This current induces a field intensity

$$H_f=\frac{0.4\,\pi i_f}{L}=-0.8\pi B_s\gamma z\,\frac{dz}{dt}10^{-8}. \tag{2}$$

During the entire process of switching the core, for the center of the laminar cross section ($z = d/2$), in which the induction still has not varied, we have

$$H+H_f=H_c \tag{3}$$

where

$$H=\frac{0.4\,\pi wi}{l}. \tag{4}$$

By taking into account the fact that the core flux Φ is related to the quantity z by the formula

$$\Phi = -\Phi_s + \frac{4z}{d}\Phi_s \quad \text{or} \quad z = \frac{\Phi + \Phi_s}{4\Phi_s} d,$$

we obtain from Equations (1-4),

$$M(\Phi + \Phi_s)\,d\Phi + N d\Phi = (h - H_c)\,dt, \tag{5}$$

where

$$M = \frac{0.8\pi B_s \gamma d^2}{16\times 10^8 \Phi_s^2}, \tag{6}$$

$$N = \frac{0.4\pi w^2}{Rl \cdot 10^8} \tag{7}$$

and

$$h = \frac{0.4\pi w}{Rl} e. \tag{8}$$

The core begins to be switched only after h, in changing from zero, attains the value H_C at time $t = t'$ $[h(t') = H_C]$. Thus, integrating Equation (5) from t' to t, which corresponds to the flux change from $-\Phi_s$ to Φ, we obtain

$$\frac{M}{2}(\Phi + \Phi_s)^2 + N(\Phi + \Phi_s) = \int_{t'}^{t} h dt - H_c(t - t') \tag{9}$$

or, using the fact that at all times $\Phi + \Phi_s \geq 0$:

$$\Phi + \Phi_s = -\frac{N}{M} + \sqrt{\frac{N^2}{M^2} + \frac{2}{M}\left[\int_{t'}^{t} h dt - H_c(t - t')\right]}. \tag{10}$$

Substituting Expression (10) for the flux in (1), and solving the resulting equation for the current, we find

$$i = \frac{e}{R} - \frac{w(h - H_c)}{RM \cdot 10^8 \sqrt{\frac{N^2}{M^2} + \frac{2}{M}\left[\int_{t'}^{t} h dt - H_c(t - t')\right]}}.$$

For the field intensity (4), we have

$$H = h - \frac{\frac{N}{M}(h - H_c)}{\sqrt{\frac{N^2}{M^2} + \frac{2}{M}\left[\int_{t'}^{t} h dt - H_c(t - t')\right]}}. \tag{11}$$

From (10) and (11) we obtain the equation for the ascending arms of the dynamic hysteresis loop:

$$H = \frac{H_c + \frac{M}{N}(\Phi + \Phi_s)\,h}{1 + \frac{M}{N}(\Phi + \Phi_s)}, \tag{12}$$

or

$$H=\frac{H_c+\frac{\gamma d^2Rl}{8Sw^2}\left(1+\frac{B}{B_s}\right)h}{1+\frac{\gamma d^2Rl}{8Sw^2}\left(1+\frac{B}{B_s}\right)},\tag{12'}$$

where S is the core cross sectional area, and $B = \Phi/S$ is the average or apparent induction.

As might be expected for $\gamma = 0$ (which occures with ferrites), and also for $d \to 0$, we get $H = H_c$, and the dynamic and static hysteresis loops coincide.

If, in Formulas (12) and (12'), we set, respectively, Φ and B equal to zero, we then obtain the value of the coercive force for the dynamic hysteresis loop:

$$H_{cD}=\frac{H_c+\frac{M}{N}\Phi_s h}{1+\frac{M}{N}\Phi_s}.\tag{13}$$

For $h = \text{const.} = H_c$, we see from (13) that $H_{cD} = H_c$.

We now consider certain examples in which the formulas so far obtained are applied.

<u>Sinusoidally Varying Induction</u>

Let $e = E_m \sin \omega t$. The induction will vary sinusoidally for $R = 0$. Integrating Equation (1) for this case, and bearing in mind that at $t = 0$ the induction $B = -B_s$, we obtain

$$B=-B_s+\frac{E_m}{\omega wS\cdot 10^{-8}}(1-\cos\omega t).\tag{14}$$

We set

$$\frac{E_m}{\omega w\cdot S10^{-8}}=kB_s.\tag{15}$$

For $R=0, \frac{1}{N}=0,$ and $\frac{h}{N}=\frac{e}{w\cdot 10^{-8}}=\frac{E_m}{w}10^8\sin\omega t.$

Therefore we have from (12), using (15):

$$H=H_c+\Phi_s^2\left(1+\frac{B}{B_s}\right)k\omega M\sin\omega t.\tag{16}$$

From (14) and (15) it is not difficult to obtain

$$\sin\omega t=\sqrt{\frac{2}{k}\left(1+\frac{B}{B_s}\right)-\frac{1}{k^2}\left(1+\frac{B}{B_s}\right)^2},$$

and Equation (16) for the dynamic hysteresis loop takes the following form:

$$H=H_c+\omega M\Phi_s^2\left(1+\frac{B}{B_s}\right)\sqrt{2k\left(1+\frac{B}{B_s}\right)-\left(1+\frac{B}{B_s}\right)^2}.\tag{17}$$

Setting $B = 0$ in (17), we obtain the following representation for the dynamic coercive force:

$$H_{cD}=H_c+\omega M\Phi_s^2\sqrt{2k-1}=H_c+f\gamma d^2B_s\cdot 10^{-8}\sqrt{2k-1}.\tag{18}$$

Fig. 4 shows the construction of the dynamic hysteresis loop from Formula (17) for the values $k = 1$ and $H_{cD} = 3H_c$. The dotted lines indicate the static loop. It is obvious from Formula (17) and Fig. 4 that the field intensity H is a nonlinear function of B during the core switching process and, consequently, that the differential permeability is a variable quantity. From (17) the following expression for the differential permeability is easily obtained:

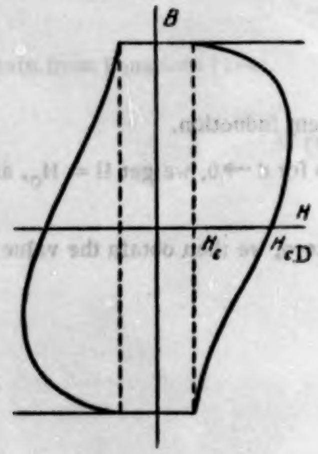


Fig. 4

$$\mu_D = \frac{dB}{dH} = \frac{B_s}{\omega M \Phi_s^2} \frac{\sqrt{2k\left(1+\frac{B}{B_s}\right)-\left(1+\frac{B}{B_s}\right)^2}}{3k\left(1+\frac{B}{B_s}\right)-2\left(1+\frac{B}{B_s}\right)^2}.$$

For $B = (\frac{3}{2}k - 1)B_s$, $\mu_D = \infty$, and the field intensity H attains its maximum value, equal to

$$H_m = H_c + \frac{3\sqrt{3}}{4} k^2 \omega M \Phi_s^2.$$

For further increase in B, the field intensity H is decreased, and μ_D assumes negative values. When $k = 1$, corresponding to the induction varying from $-B_s$ to $+B_s$ during a half-period of the supply voltage, H attains its maximum, and $\mu_D = \infty$ for $B = +0.5\,B_s$ (Fig. 4). If $k > 4/3$, then H does not attain its maximum, since this would have been observed at $B > B_s$.

Sinusoidally Varying Field Intensity

We let

$$h = \frac{0.4\pi w E_m}{Rl} \sin \omega t = H_m \sin \omega t.$$

We obtain the same law of variation for H if $R \to \infty$, but the ratio $E_m/R = \text{const}$. In this case $N \to 0$ and Equation (10) assumes the following form:

$$\Phi + \Phi_s = \sqrt{\frac{2}{\omega M}[H_m(\cos\omega t' - \cos\omega t) - \omega H_c(t - t')]}, \tag{19}$$

and, from (12), we obtain

$$H = h = H_m \sin \omega t. \tag{20}$$

Since $H_m \sin \omega t' = H_c$, we can get rid of t and t' in Fomula (19), obtaining for the dynamic hysteresis loop

$$B = -B_s + \frac{1}{S}\sqrt{\frac{2}{\omega M}\left[H_m\left(\sqrt{1-\frac{H_c^2}{H_m^2}} - \sqrt{1-\frac{H^2}{H_m^2}}\right) - H_c\left(\arcsin\frac{H}{H_m} - \arcsin\frac{H_c}{H_m}\right)\right]} \tag{21}$$

The case when $H_m \gg H_c$ is particularly interesting.

Since, in this case, the switching of the core occurs for a value $H \ll H_m$, we may substitute in Formula (21)

$$\sqrt{1-\left(\frac{H_c}{H_m}\right)^2} \approx 1 - \frac{1}{2}\left(\frac{H_c}{H_m}\right)^2, \quad \sqrt{1-\left(\frac{H}{H_m}\right)^2} \approx 1 - \frac{1}{2}\left(\frac{H}{H_m}\right)^2,$$
$$\arcsin\frac{H_c}{H_m} \approx \frac{H_c}{H_m} \text{ and } \arcsin\frac{H}{H_m} \approx \frac{H}{H_m}. \tag{22}$$

Making these substitutions in (21), we find *

$$B = -B_s + \frac{H - H_c}{S\sqrt{\omega M H_m}}. \qquad (22)$$

Thus, for $H_m \gg H$, and also for a linear variation of H in time, the induction B varies quite proportionately to H, but the differential permeability remains essentially constant during the entire process of core switching, and equals

$$\mu_D = \frac{dB}{dH} = \frac{1}{S\sqrt{\omega M H_m}} = 10^4\sqrt{\frac{B_s}{f\gamma d^2 H_m}}. \qquad (23)$$

It might be mentioned that μ_D is frequently measured by oscillographic methods in order to determine the number of cores to be used in magnetic amplifiers [1]. For this purpose, the cores are magnetized by the sinusoidal field (20), for which $H_m \gg H_c$. Then, the sine occuring in Formula (20) may be replaced by its argument (20'), and for the emf induced i.. core winding w_2 we obtain

$$e = -wS10^{-8}\frac{dB}{dt} = -wS10^{-8}\frac{dB}{dH}\frac{dH}{dt} = -\omega w S H_m 10^{-8}\mu_D$$

It is clear from Formula (23) that the variation of μ_D depends essentially on the quantity H_m, which must be accounted for in carrying out the measurements.

If we set $B = 0$ in Formula (22), we obtain the following expression for the magnitude of the dynamic coercive force when $H_m \gg H_{cD}$:

$$H_{cD} = H_c + \Phi_s\sqrt{\omega M H_m} = H_c + \sqrt{10^{-8} f B_s \gamma d^2 H_m}. \qquad (24)$$

DC Magnetization

Let $e = E = \text{const}$ and, correspondingly, $h = H_0 = \text{const}$. In this case the dynamic hysteresis loop is determined from Formula (12) or from Formula (12'), and the dynamic coercive force, after substituting H_0 for h, from Formula (13). It is easily seen that for $H_0 = H_c$, we have $H_{cD} = H_c$.

In many types of magnetic amplifiers with positive feedback, the magnitude of the load voltage is determined by the value of the induction B_y which is established in the core during a half-period (T/2) of the supply voltage, when acted upon by a dc signal (cf., for example, [1], par. 7-5). Taking into account that $t' = 0$ in the case under consideration, and substituting $t = T/2 = 1/2f$, we find from Formula (10)

$$B_y = -B_s - \frac{N}{MS} + \frac{1}{S}\sqrt{\frac{N^2}{M^2} + \frac{1}{fM}(H_0 - H_c)}. \qquad (25)$$

The gain, either of voltage or of current, is proportional to the derivative [1] :

$$\frac{dB_y}{dH_0} = \frac{1}{2S\sqrt{f^2N^2 + fM(H_0 - H_c)}}. \qquad (26)$$

In the absence of eddy currents, $M = 0$, and the derivative dB_y/dH_0 does not depend on the magnitude of

* It might be mentioned that the condition $H \ll H_m$ corresponds to a linear variation in time of the field intensity

$$H = h = H_m\omega t \qquad (20')$$

and Formula (22) may be obtained directly from Equations (10) and (20').

the signal, H_0 (or E). The presence of eddy currents not only decreases the gain, but also makes it a function of the signal. An increase of the latter causes the derivative, and hence the gain, to decrease.

If in Formula (25) we set $B_y = B_s$, we obtain the following expression for $H_{0\,max}$, the field intensity of the signal necessary to magnetize a core completely in one half-period of supply voltage:

$$H_{0\,max} = H_c + 4fMS^2B_s^2 + 4fNSB_s.$$

Thus, the presence of eddy currents entails the necessity of increasing the field intensity of the signal by the quantity

$$\Delta H_0 = 4fMS^2B_s^2 = 2\pi\cdot 10^{-9}f\gamma d^2 B_s.$$

It is also possible to express ΔH_0 by means of the dynamic coercive force under sinusoidal induction. Indeed, taking into account Equation (18), we find that for $k = 1$

$$\Delta H_0 = 0.2\pi\,(H_{cD} - H_c).$$

Thus, if H_c is known, it is sufficient to measure H_{cD} under sinusoidal induction for evaluating the influence of eddy currents on the behavior of a magnetic amplifier.

Influence of Short-Circuited Windings and Measuring Circuits on the Experimental Determination of Dynamic Characteristics

In certain cases, the presence of shorted core windings, and also the presence of finite resistance in the measuring device, can significantly distort the results in measuring the parameters of a dynamic hysteresis loop.

We suppose that around the core (in addition to the basic magnetizing winding, w) is a second winding w_k, shorted by a resistance R_k. Then, on the left side of Equation (3), it is necessary to append the field intensity H_k induced by the current i_k in winding w_k. For H_k we have

$$H_k = \frac{0.4\pi w_k}{l}\, i_k = -\frac{0.4\pi w_k}{l}\,\frac{w_k}{R_k\cdot 10^8}\,\frac{d\Phi}{dt}. \tag{27}$$

Letting

$$K = \frac{0.4\pi w_k^2}{R_k l\cdot 10^8},$$

we obtain, instead of (9), the following expression:

$$\frac{M}{2}(\Phi + \Phi_s)^2 + (N + K)(\Phi + \Phi_s) = \int_{t'}^{t} h\,dt - H_c(t - t'), \tag{28}$$

and the equation of the apparent dynamic hysteresis loop takes the following form:

$$H = \frac{H_c + \frac{K}{N}h + \frac{M}{SN}(B + B_s)h}{1 + \frac{K}{N} + \frac{M}{SN}(B + B_s)}. \tag{29}$$

It may be concluded from (29) that shorted windings have a completely different effect on the form of the dynamic hysteresis loop than do eddy currents. Therefore, in calculating the influence of the latter, no recourse was had to introducing "equivalent" shorted windings, as is sometimes done.

If the induction is varying sinusoidally (14), it is possible to obtain the following expression for the

apparent dynamic hysteresis loop, using the previously investigated method:

$$H = H_c + \omega\Phi_s\left[K + M\Phi_s\left(1 + \frac{B}{B_s}\right)\right]\sqrt{2k\left(1 + \frac{B}{B_s}\right) - \left(1 + \frac{B}{B_s}\right)^2}. \quad (30)$$

Letting $B = 0$, we find from (30) that the presence of a winding w_k shorted by a finite resistance, leads to an increase in the apparent dynamic coercive force by the amount

$$\Delta H_{ck} = \omega\Phi_s K\sqrt{2k - 1}. \quad (31)$$

EXPERIMENTAL

In a work by the author [4] it was shown that the method herein employed for studying the influence of eddy currents permits of sufficient accuracy in computing the process of magnetizing strip cores of alloy N34K29MZP with rectangular static hysteresis loops. Here, we cite certain results of investigations of the dynamic characteristics of toroidal cores made from alloy 65NP. The cores were made up of rings 0.35 mm thick, 30 mm outside diameter and 20 mm inside diameter. The specific electrical conductivity of the alloy equals 4×10^4 (ohm cm)$^{-1}$, and the static coercive force of the cores was $H_c = 0.06$ oersted. In evaluation of the formulas, instead of B_s, the value of the residual induction B_r was used; this value turned out to be 12,000 gauss.

Fig. 5

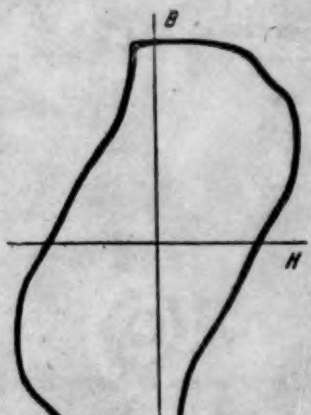


Fig. 6

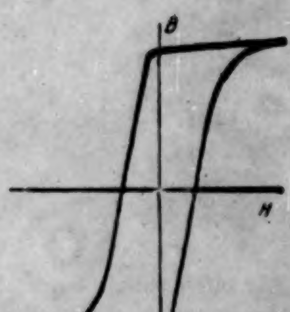


Fig. 7

Figures 5 and 6 give oscillograms of the dynamic hysteresis loops, taken off, respectively, for frequencies of 50 and 500 cps, with sinusoidally varying induction ($k = 1$). It is clear that the general, somewhat unusual, character of the dependence $B = f(H)$ corresponds to the theoretical curve of Fig. 4. The oscillogram of Fig. 7, obtained with a sinusoidally varying field intensity, shows that in this case the dependence $B = f(H)$ is virtually linear during the course of practically the entire magnetizing processes of the core, in accordance with Formula (22).

Table 1 contains the results of computations using Formula (18), and the results of measuring the dynamic coercive force H_{cD} for various values of $\underline{k}$ and of frequency $\underline{f}$ with a sinusoidally varying induction.* Table 2 gives the results of computation based on Formula (24) and the results of measuring H_{cD} with sinusoidally varying field intensity.

In Table 3 are given the results of evaluating Formula (23) and the results of meauring the differential permeability μ_D for a field intensity varying sinusoidally with a frequency of 50 cps.

The results given in Table 1-3 show a satisfactory agreement between the computed and the measured values

* Strictly speaking, induction varies sinusoidally only for $k \le 1$. For > 1, the induction varies from $-B_s$ to $+B_s$ according to the cosine law, remains invariant during a certain time interval, and then varies again according to the cosine law from $+B_s$ to $-B_s$.

TABLE 1

k	f, cps	H_{cD}, oersted measured	H_{cD}, oersted computed from (18)
1	50	0.36	0.355
2	50	0.63	0.57
3	50	0.83	0.72
4	50	0.95	0.84
5	50	1.08	0.94
6	50	1.20	1.04
1	500	2.97	3.01
2	500	4.85	5.12

of H_{cD} and μ_D; however, it should be borne in mind that, in measuring these quantities by the oscillogram method, one obtains an accuracy no better than 5-10 %.

The same agreement between theory and experiment was also obtained with cores prepared from 0.133 mm thick 65NP alloy. For cores of material 0.05 and 0.02 mm thick, satisfactory agreement between computed and measured characteristics was obtained only with magnetization by sinusoidally varying, or by constant, fields, with H_m or $H_0 \gg H_c$. For values of H_m very close to H_c, and with sinusoidal induction, the measured values of H_{cD} turned out to be significantly higher than the computed ones, which might have resulted either from the influence of magnetic viscosity, or from the inhomogeneity of magnetic properties along the cross section of the material.

TABLE 2

H_m, oersted	f, cps	H_{cD}, oersted measured	H_{cD}, oersted computed from (24)
2	50	0.75	0.82
4	50	1.07	1.14
6	50	1.46	1.38
8	50	1.62	1.56
10	50	1.82	1.77
12	50	2.06	1.94

TABLE 3

H_m, oersted	μ_D, gauss/oersted computed from (23)	μ_D, gauss/oersted measured
2	15 700	17 000
4	11 100	11 800
6	9 050	8 600
8	7 850	7 700
10	7 000	6 800
12	6 400	6 000

SUMMARY

1. The influence of eddy currents on the processes of magnetizing cores with rectangular static hysteresis loops may be studied by a method due to V. K. Arkad'ev.

2. The form of the dynamic hysteresis loop and the magnitude of the dynamic coercive force depend, in an essential fashion, on the law of variation of the magnetic induction or of the field intensity.

3. With sinusoidally varying induction it is possible to observe a peculiar form of the dynamic hysteresis loop, characterized by the following behavior: as the induction increases, the field intensity increases to a certain maximum value, at which the differential permeability becomes infinite, then decreases again. This allows the differential permeability to assume negative values.

4. Eddy currents cause a lowering of the gain and a disturbance of the linearity characteristics of magnetic amplifiers with positive feedback, controlled by dc signals. The influence of eddy currents may be estimated by measuring the coercive force with direct current, and sinusoidally varying induction.

5. The effect of shorted windings is different from that of eddy currents. It is impossible to simulate the effect of eddy currents on the dynamic hysteresis loop by the introduction of "equivalent" shorted windings.

6. A significant influence of eddy currents of dynamic hysteresis loops is disclosed at frequencies as low as 50 cps, even for core materials 0.05-0.1 mm thick, and for material thinner than that, particularly with a sinusoidally varying field intensity.

LITERATURE CITED

[1] M. A. Rozenblat, Magnetic Amplifiers [in Russian] (Izd. "Sovetskoe Radio," 1956).

[2] N. Menyuk, "Magnetic Materials for Digital Computer Components. Part 2, Magnetic characteristics of ultra-thin molybdenum permalloy cores," J. Appl. Phys. 26, 6 (1955).

[3] V. K. Arkad'ev, "Two methods of computing skin effect in ferromagnetics" [in Russian] Coll.: Practical Problems of Electromagnetics (Izd. AN SSSR, 1939).

[4] M. A. Rozenblat, "Pulse magnetized cores with rectangular hysteresis loops," Elektrichestvo 1955, No. 1.

Received May 15, 1957

SOME OPTIMAL RELATIONSHIPS IN IDEAL AC-CONTROLLED MAGNETIC AMPLIFIERS*

K. S. Volchkov

(Moscow)

The functioning of ideal ac-controlled, active-load, choked magnetic amplifiers is investigated. Optimal relationships are deduced for amplifiers designed either for amplifying signals of one fixed frequency or for amplifying ac signals in some given frequency band; such relationships guarantee maximum gain. The connection is established between an amplifier's frequency distortion coefficient and its time constant.

At the present time there is a widespread use of semiconductor amplifiers without vacuum tubes for the amplification of ac signals of audible and higher frequencies. However, for ac amplification, it is sometimes more suitable to use not crystal, but magnetic amplifiers, since the latter, in certain respects, possess more favorable properties. Thus, for example, magnetic amplifiers have lower excitation thresholds than do crystal ones, they are more easily impedance matched with arbitrary low-resistance transducers and receivers, allow more successful amplification of very low frequencies, and do not have limitations on their output power. Moreover, magnetic amplifiers show little variation in parameters from one replicate to another, and their characteristics are quite stable under variations in temperature. Considering that the properties of crystal and magnetic amplifiers frequently complement one another, it sometimes appears reasonable to use the two types in conjunction.

The present paper is devoted to an exposition of the optimal relationships for ac-controlled magnetic amplifiers. Herein are submitted to analysis ideal choking amplifiers with active loads, whose ac output signals are transformed into the modulating vibrations of a carrier frequency. If two half-period rectifiers are used, this frequency is twice that of the amplifier supply. In the sequel, such amplifiers will be called ac amplifiers.

Let the transducer signals entering the magnetic amplifier input (Fig. 1) induce an emf, the angular frequency of which is Ω, with amplitude E_{Dm}. (For simplicity we shall assume that the signal transducer has only an active resistance, R_D, although analogous reasoning could be followed even for a transducer with complex impedance.)

In the control circuit of the amplifier will flow a current, the instantaneous value of which is determined from the following formula [1]:

$$i_y = \frac{E_{Dm}}{R_y(1+\alpha)\sqrt{1+\Omega^2\tau^2}} \sin(\Omega t - \varphi), \tag{1}$$

where $\alpha = R_D/R_y$ is a coefficient characterizing the ratio between the transducer's active resistance and the control winding resistance, and τ is the time constant of the control circuit.

Within the limits of the operating region of its characteristics, $I_H = f(I_y)$ (where I_H is the load circuit current), the current gain of an ideal amplifier, and its time constant τ depend neither on the magnitude nor on the frequency of the signal. Consequently, there will correspond to the control current i_y a completely determinate load current,** the variable component of which may be found from the current gain K_I by the use of

* Pesented at a seminar on magnetic amplifiers at IAT AN SSSR, Feb. 13, 1957.

** This take the form of the average current value during a half-period of supply voltage.

the following equation:

$$i_H = \frac{K_i E_{Dm}}{R_y(1+\alpha)\sqrt{1+\Omega^2\tau^2}}\sin(\Omega t - \varphi).$$

The effective value of the power thus dissipated by the active load resistance R_H will equal

$$P_H = I_H^2 P_H = K_i^2 I_y^2 R_H.$$

Substituting here the effective value I_y from Expression (1), and the known representation for K_i [1], we obtain

$$P_H = \frac{W_y^2 R_H}{W_\sim^2 R_y(1-K_{fb})^2}\,\frac{E_D^2}{R_y(1+\alpha)^2}\,\frac{1}{1+\Omega^2\tau^2}, \quad (2)$$

where W_y and $W_\sim$ are the number of turns in the control and ac current windings, and K_{fb} is the feedback coefficient.

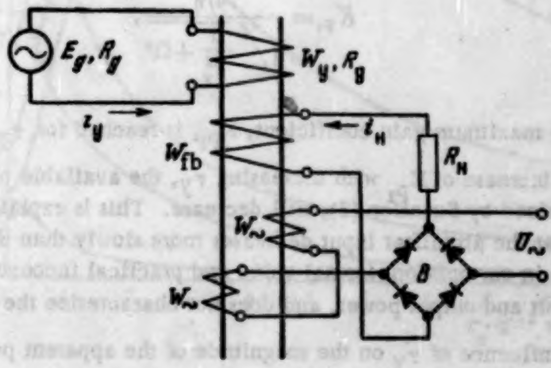


Fig. 1. Circuit connecting amplifier and signal transducer.

Taking into account that the first factor in Expression (2) is the power gain of the amplifier at dc, we rewrite (2) in the following form:

$$P_H = \frac{4f\eta}{1-K_{fb}}\,\frac{E_D^2}{R_y(1+\alpha)^2}\,\frac{\tau_y}{1+\Omega^2\tau^2}, \quad (3)$$

where f is the supply voltage frequency, and η is the efficiency of the magnetic amplifier.

Expressing the time constant of the control circuit, τ in terms of the time constant τ_y of the amplifier itself

$$\tau = \frac{\tau_y}{1+\alpha} \quad (4)$$

and substituting this in (3), we get

$$P_H = \frac{4f\eta}{1-K_{fb}}\,\frac{E_D^2}{R_y}\,\frac{\tau_y}{(1+\alpha)^2+\Omega^2\tau_y^2}. \quad (5)$$

We find the optimal value of τ_y (given Ω, R_y, α, etc.), for which P_H attains a maximum. Taking the derivative, $\partial P_H/\partial\tau_y$, and equating it to zero, we find

$$\tau_{y\,opt} = \frac{1+\alpha}{\Omega}. \quad (6)$$

Using (4), this relationship may be put in a simpler form:

$$\tau_{opt} = \frac{1}{\Omega}. \tag{7}$$

It follows from this equation that increasing the frequency entails choosing a lower value for the time constant.

In evaluating and comparing magnetic amplifiers, it is convenient to use the power gain coefficient. However, for ac magnetic amplifiers the power gain coefficient may be expressed in different ways, and an established method of defining it does not exist. Certain authors understand by "power gain coefficient" the ratio of the amplifier output power to the apparent input power:

$$K_{p_1} = \frac{(\Delta I_н)^2 R_н}{(\Delta I_y)^2 R_y \sqrt{1 + \Omega^2 \tau_y^2}}.$$

If the given equation takes the form

$$K_{p_1} = \frac{4/\eta}{\sqrt{\frac{1}{\tau_y^2} + \Omega^2}}, \tag{8}$$

then it may be seen that the maximum gain coefficient, K_{p_1}, is reached for $\tau_y \rightarrow \infty$.

Despite the continous increase of K_{p_1} with increasing τ_y, the available power in the load, beyond a certain optimal value of τ_y, determined by Equation (6), will decrease. This is explained by the fact that, with increasing τ_y, the apparent power at the amplifier input decreases more slowly than the output power. It is for this reason that the coefficient K_{p_1} has, in our opinion, formal value and practical inconvenience, since it indicates only the relationship of amplifier input and output power, and does not characterize the available power at the load.

We now consider the influence of τ_y on the magnitude of the apparent power at the amplifier input. From the expression for the apparent power

$$S_y = I_y^2 Z_y = \frac{E_D^2 \alpha \sqrt{1 + \Omega^2 \tau_y^2}}{R_D [(1+\alpha)^2 + \Omega^2 \tau_y^2]} \tag{9}$$

it can be found that, if $\alpha + 1 < \sqrt{2}$, then the maximum induced power in the control winding occurs at $\tau_y = 0$; if $\alpha + 1 > \sqrt{2}$, the maximum power is attained with time constant

$$\tau_y = \frac{1}{\Omega} \sqrt{(1+\alpha)^2 - 2}. \tag{10}$$

A graphic representation of the function $S_y = f(\tau_y)$ for $\Omega = 316$ is given in Fig. 2. Obviously, the function $S_y = f(\Omega)$ will have the same form.

Since it is clear that $\alpha + 1 > \sqrt{2}$, the transducer signals are computed necessarily on the basis of maximum amplifier output, obtained by satisfying Condition (10) and determined from the expression

$$S_{y\,max} = \frac{E_D^2}{2R_D} \frac{1}{\sqrt{1 + \frac{2}{\alpha}}}.$$

Comparing Expression (6) and (10), one may see that after the value τ_y, determined by Equation (10), is attained, the magnitude of the apparent power at the amplifier input will fall, while the power in the load winding will continue to increase until such time as τ_y attains the value determined by Relationship (6).

The power gain coefficient may also be expressed as the ratio of output power to the active power dissipated in the control winding:

$$K_{p2} = \left(\frac{\Delta I_н}{\Delta I_y}\right)^2 \frac{R_н}{R_y} = K_{p0}. \quad (11)$$

But in this case, the ac gain K_{p_2} turns out to equal the dc gain K_0 and, just as for K_{p_1}, does not characterize the amplifier output power, since the admission of active power at its input depends on the signal frequency.

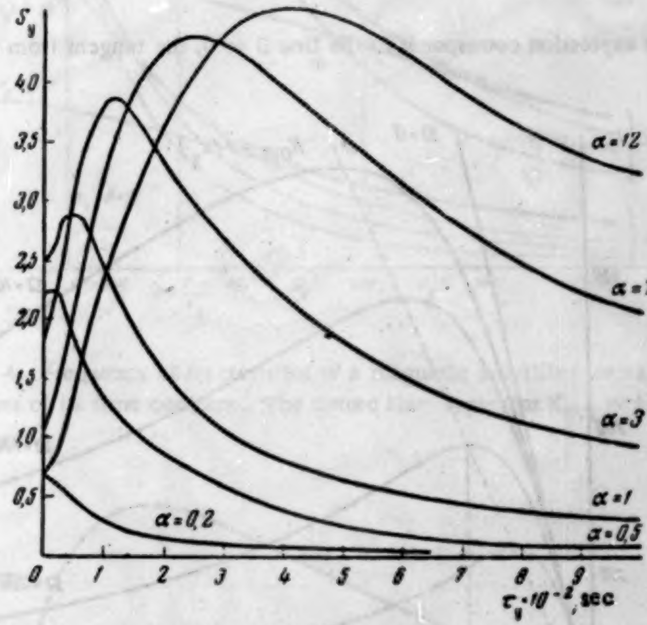


Fig. 2. Dependence of a magnetic amplifier's apparent input power on its time constant.

It is always satisfactory to define the power gain coefficient as the ratio of the effective power produced in the load to the maximum power which the signal transducer is capable of providing. Such a definition is advantageous in this respect, that, by means of the gain coefficient, one may judge the increase of effective load power, and the degree of utilization of all amplifying devices in units comprised also of signal transducers. Moreover, it appears to be sufficiently general, since it is acceptable for arbitrary amplifier types, including those amplifiers for which the signal transducers are not generators, but require energy (e.g., Ramey magnetic amplifiers [2]).

The maximum power which a transducer is potentially capable of providing will equal

$$P_D = \frac{E_D^2}{4R_D}. \quad (12)$$

Dividing the left and right member of Equation (5) by the corresponding members of (12), and replacing R_y in the latter by R_D/α, we obtain an expression for the power gain coefficient:

$$K_{p_3} = A \frac{\alpha \tau_y}{(1+\alpha)^2 + \Omega^2 \tau_y^2}, \quad (13)$$

where $A = \frac{16 f \eta}{1 - K_{fb}}$. Fig. 3 give graphs of Function (13) as a function of τ_y for different values of Ω, and with $A = 10^5$ and $\alpha = 3$. It is apparent from the curves that K_{p_3} has a clearly defined maximum which occurs when Condition (6) is satisfied. The greatest possible gain coefficient for the given signal frequency will be called the

optimal power gain coefficient, K_{opt}. Substituting the value of Ω from (6) in Equation (13), we find

$$K_{opt} = \frac{8/\eta}{1 - K_{fb}} \frac{\alpha\tau_y}{(1+\alpha)^2}. \tag{14}$$

In accordance with Equation (14), all the optimal gain coefficients on Fig. 3 fall on the straight line OA. If, in Equation (13), we set $\Omega = 0$, we then obtain the expression for the dc gain coefficient:

$$K_{p_0} = \frac{16/\eta}{1 - K_{fb}} \frac{\alpha\tau_y}{(1+\alpha)^2} = \frac{4K_{p0}\alpha}{(1+\alpha)^2}. \tag{15}$$

In Fig. 3, this last expression corresponds to the line $\Omega = 0$, the tangent from the origin to the family of curves $K_{p_3} = f(\tau_y)$.

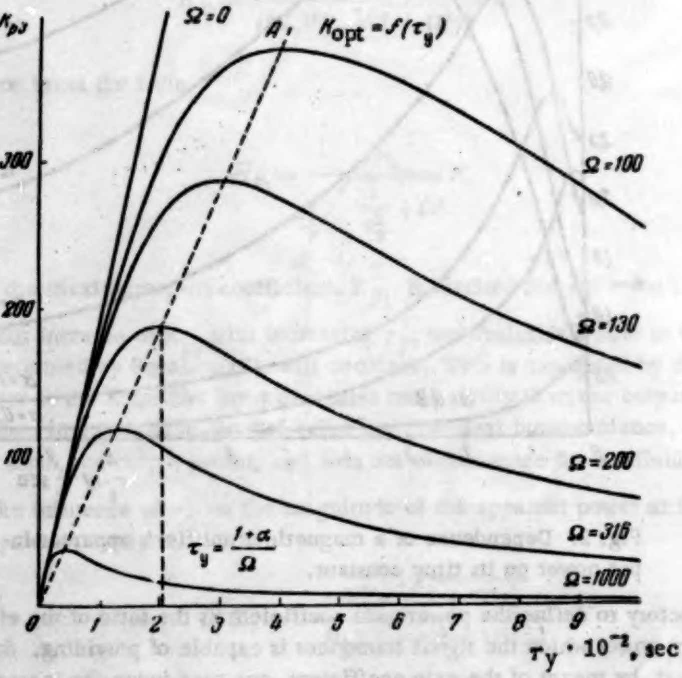


Fig. 3. Dependence of power gain coefficient K_{p_3} on the magnetic amplifier's time constant, for various values of input signal frequency.

Fig. 4 represents the same Function (13) as it depends on Ω, with τ_y as a parameter, in the case where $\alpha = 5$. The graphs show two families of curves, corresponding to two values of the coefficient A in Equation (13).

In order to satisfy Condition (6), Equation (13) is transformed to:

$$K_{p3} = K_{opt} = \frac{8/\eta}{1 - K_{fb}} \frac{\alpha}{1+\alpha} \frac{1}{\Omega}, \tag{16}$$

from which it is easy to see that the geometric locus of the optimal value of K_{p_3} for variable frequency Ω is a hyperbola. In Fig. 4, these hyperbolas are represented by dotted lines for each family of curves. The curve $K_{p_3} = f(\Omega)$ is then the hyperbola made up of the points for which Equation (6) is satisfied.

The gain coefficient K_{p_3}, expressed by Equation (13), will have extremal values for variable α and fixed τ_y (Fig. 5). To determine the values of α for which K_{p_3} attains a maximum, we set the derivative $\partial K_{p_3}/\partial \alpha$ equal to zero, after which we find that

$$\alpha = \sqrt{1 + \Omega^2\tau_y^2}. \tag{17}$$

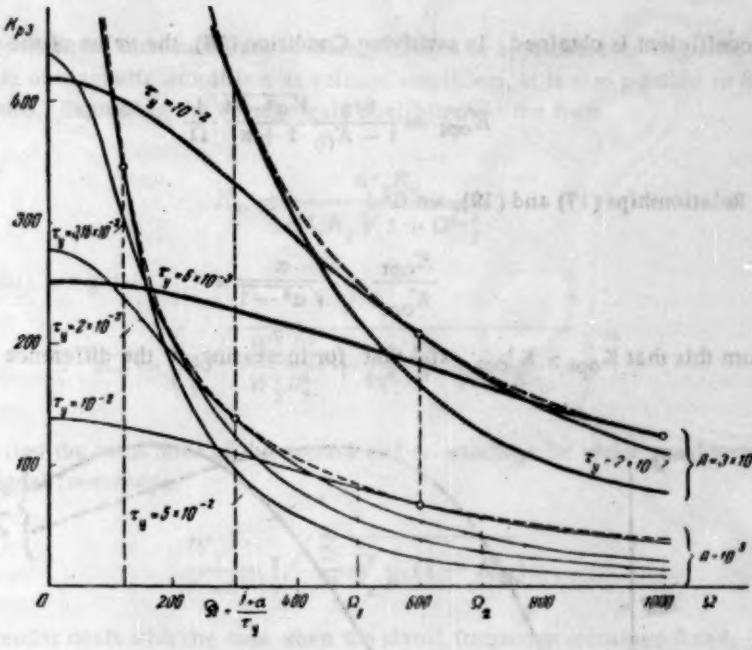


Fig. 4. Frequency characteristics of a magnetic amplifier for various values of its time constant. The dotted lines represent $K_{opt} = f(\Omega)$.

Fig. 5. Dependence of the gain coefficient K_{p_3} on the ratio of the impedances of the transducer and of the amplifier control winding ($\alpha = R_D/R_y$) for $\tau_y = 10^{-2}$ and $A = 10^5$

Equation (17) essentially expresses the condition that the transducer impedance be matched to the magnetic amplifier input impedance Z_y since, in this case,

$$R_D = R_y\sqrt{1+\Omega^2\tau_y^2} = Z_y. \tag{18}$$

It is necessary to determine from which of the two mutually exclusive conditions, (6) and (18), the greatest

value of the gain coefficient is obtained. In satisfying Condition (18), the value of the gain coefficient will equal

$$K'_{opt} = \frac{8/\eta}{1 - K_{fb}} \frac{\sqrt{\alpha^2 - 1}}{1 + \alpha} \frac{1}{\Omega}. \tag{19}$$

Considering Relationships (17) and (19), we find

$$\frac{K_{opt}}{K'_{opt}} = \frac{\alpha}{\sqrt{\alpha^2 - 1}}. \tag{20}$$

It follows from this that $K_{opt} > K'_{opt}$, and that for increasing α the difference between them decreases.

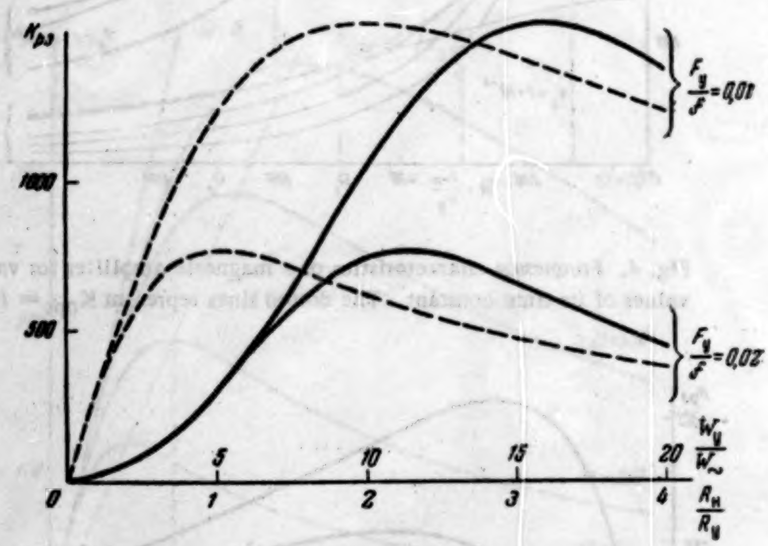


Fig. 6. Dependence of the gain coefficient K_{p_3} on the magnetic amplifier parameters $W_y/W_\sim$ (solid line) and R_H/R_y (dotted line).

The condition introduced above, (7), for obtaining the optimal power gain coefficient allows a rational choice of parameters for an ac magnetic amplifier. If we substitute in Equation (7) the value

$$\tau = \frac{W_y^2 R_H}{W_\sim^2 R_y (1 + \alpha) 4/\eta (1 - K_{fb})}, \tag{21}$$

then we can get the expression

$$\frac{\pi}{2\eta} \frac{F}{f} \frac{W_y^2}{W_\sim^2} \frac{R_H}{R_y} \frac{1}{(1 - K_{fb})(1 + \alpha)} = 1, \tag{22}$$

establishing the connection between the basic amplifier parameters necessary for obtaining optimal amplification for the given signal frequency, $F = \Omega/2\pi$. For example, using the charging coefficients K_{3y} and $K_{3\sim}$ of the control and ac windings, plus the cross sectional areas of their conductors, q_y and $q_\sim$, we can find from Equation (22) the required ratio of their aperture areas:

$$\frac{Q_\sim}{Q_y} = \frac{\pi}{2\eta} \frac{R_H}{R_y} \frac{W_y}{W_\sim} \frac{K_{3y}}{K_{3\sim}} \frac{q_\sim}{q_y} \frac{F}{f} \frac{1}{(1 - K_{fb})(1 + \alpha)}. \tag{23}$$

Fig. 6 indicates the influence of the ratios $W_y/W_\sim$ and R_H/R_y on the gain coefficient K_{p_3} for $\alpha = 3$, $\eta = 0.8$ and $K_{fb} = 0.95$.

The functional dependences introduced above were obtained as applicable to magnetic amplifiers of power or of current. For the use of magnetic amplifiers as voltage amplifiers, it is also possible to find the most advantageous amplifier parameters. Expressing the voltage gain coefficient in the form

$$K_u = \frac{\Delta I_{\text{н}} R_{\text{н}}}{\Delta I_y R_y \sqrt{1+\Omega^2\tau_y^2}}$$

and using Relationship (21), we get

$$K_u = \left(\sqrt{\frac{W_\sim^2 R_y^2}{W_y^2 R_{\text{н}}^2} + \frac{\pi^2}{4\eta^2}\frac{F^2}{f^2}\frac{1}{(1-K_{fb})^2}}\right)^{-1}. \tag{24}$$

By using (24), we find the turns ratio of the control and ac windings for which maximum amplification is attained for the given signal frequency:

$$\frac{W_y}{W_\sim} = \sqrt{\frac{2\eta}{\pi}\frac{R_y}{R_{\text{н}}}\frac{f}{F}(1-K_{fb})}. \tag{25}$$

All the previous results dealt with the case when the signal frequency remained fixed. For the behavior of an amplifier in a certain frequency band, the computations may be carried out starting either from a given frequency distortion or from the maximum average power gain coefficient.* The average gain coefficient in the given frequency band from Ω_1 to Ω_2 (Fig. 4) is found from the following expression:

$$K_{ps\,av} = \frac{1}{\Omega_2-\Omega_1}\int_{\Omega_1}^{\Omega_2} K_{ps}(\Omega)\,d\Omega = \frac{1}{\Omega_2-\Omega_1}\int_{\Omega_1}^{\Omega_2}\frac{A\alpha\tau_y}{(1+\alpha)^2+\Omega^2\tau_y^2}\,d\Omega =$$
$$= \frac{A}{\Omega_2-\Omega_1}\frac{\alpha}{1+\alpha}\left[\arctan\frac{\tau_y\Omega_2}{1+\alpha} - \arctan\frac{\tau_y\Omega_1}{1+\alpha}\right]. \tag{26}$$

Setting the derivative $\partial K_{ps\,av}/\partial\tau_y$ equal to zero, we find the condition for which the average power gain coefficient in the given band will be a maximum:

$$\tau_y = \frac{1+\alpha}{\sqrt{\Omega_1\Omega_2}}. \tag{27}$$

For amplifier computations starting from the frequency distortion, there is usually given a coefficient of frequency distortion. Such a coefficient (for example, the voltage at frequency Ω_2 with respect to that at Ω_1) may be found from Equation (1):

$$M_u = \left|\frac{U_{\text{н}_1}}{U_{\text{н}_2}}\right| = \sqrt{\frac{1+\Omega_2^2\tau^2}{1+\Omega_1^2\tau^2}}. \tag{28}$$

Solving this equation for τ, we get

$$\tau = \sqrt{\frac{M_u^2-1}{\Omega_2^2-\Omega_1^2 M_u^2}}. \tag{29}$$

Analogously, if the power coefficient of frequency distortion is given, we have

$$\tau = \sqrt{\frac{M_p-1}{\Omega_2^2-\Omega_1^2 M_p}}. \tag{30}$$

* It is clear that in those cases for which the signal always has one and the same non-sinusoidal form, the amplifier computations must be carried out with this peculiarity taken into account.

Tests carried out on a series of magnetic amplifiers, with $K_{fb} \leq 0.95$, showed that Equations (29) and (30) are in sufficiently good agreement with experimental data. This allows one to assume that the time constants of real ac magnetic amplifiers remain practically invariant with changes in signal frequency (of the order of 1/5 of the supply frequency). The value of Formulas (29) and (30) consists not only in the possibility of computing from them the time constant, given the frequency distortion, but also in the possibility of determining τ by experimental determination of the frequency characteristics.

It is clear from Fig. 4 that in those cases where it is necessary to have the very lowest frequency distortion, it makes sense to use the upper, almost horizontal, portion of the curve $K_{p_3} = f(\Omega)$. The range of this area may be extended either at the cost of increasing the coefficient A (whereby the gain coefficient is also increased) or of lowering τ_y (with simultaneous lowering of K_{p_3}).

It should be mentioned that Conditions (6) and (27) for obtaining maximum power gain coefficients do not afford the simultaneous attainmentof minimum frequency distortion,since in this case the operative points on the curves $K_{p_3} = f(\Omega)$ would be only the point of inflection.

SUMMARY

One can not effectively determine the power gain coefficient of a magnetic amplifier destined for ac amplification by means of the apparent or operative power at the input. The existing formulas for this purpose do not allow one to judge the completeness of utilization of every amplifying device. Moreover, these formulas cannot be applied for every type of magnetic amplifier.

It was suggested that the power gain coefficient be found as the ratio of the power developed in the load to the maximum power which the signal transducer matched to the load is capable of supplying. In an ideal magnetic amplifier controlled by some sinusoidal ac signal, maximum power gain (also maximum power in the load) is attained when the time constant of the amplifier control circuit equals the reciprocal of the signal's angular frequency. This condition determines the basic parameters of the amplifier, and allows a rational computation of them. If it is desired to obtain the maximum average power gain coefficient for a magnetic amplifier working in some frequency band from Ω_1 to Ω_2, then its time constant must be set equal to $\tau = 1/\Omega_1\Omega_2$. If the amplifier's frequency distortion coefficient is given, its time constant can be determined from Equations (29) or (30).

Received January 30, 1957

LITERATURE CITED

[1] M. A. Rozenblat, Magnetic Amplifiers [in Russian], (Izd. "Sovetskoe Radio," 1956).

[2] R. A. Lipman and I. B. Neghevitskii, "On the theory of one-half-period magnetic amplifiers," Automation and Remote Control (USSR) 18, 4 (1957).

CRITIQUE

ON ESSENTIAL INADEQUACIES OF THE DEFINITIONS AND TERMINOLOGY OF AUTOMATIC CONTROL*

A. L. Vertsaizer and V. M. Iakovlev

(Daugavpils)

Certain critical remarks are presented, bearing on the essential inadequacies of the definitions, terminology, and ambiguities presented by the structure of designs for systems of automatic control.

The absence of unanimity on elements and principles, ambiguities and multiple meanings in the presentation of designs, in the definitions and in the terminology of automatic control systems—all these are frequently to be found in contemporary literature, and particularly in textbooks, on the subject of automatic control.

Let us turn to the textbooks on automatic control which have appeared during the last decade. The textbook authors named below adhere to a single system of definitions for automatic control systems, and for the nomenclature of automatic controllers. However, the schematics of automatic control systems are given variously, not only in the sense of the sketches, but also, in our opinion, in the principles themselves.

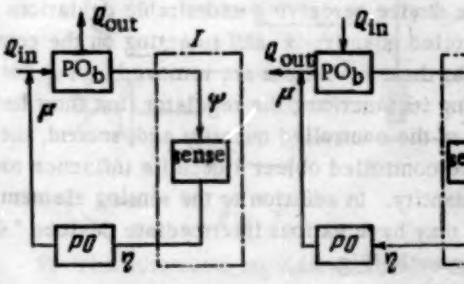


Fig. 1. I is the automatic controller

Consider book [1]. We read on page 19 that: "The controlled object, the regulating organ and the automatic controller are the elements of an automatic control system. The automatic controller and the regulating organ together comprise the regulating device by means of which the controlling process is implemented. Certain authors consider the controlling organ as an element of the automatic controller. It appears to us that this is not completely correct, since the controlling organ might also be set into motion by manual means."

Later, on page 32, we read: "A device which is used for the automatic displacement of a regulating organ in order to obtain a previously given nominal value of the controlled parameter is called an automatic controller." In Fig. 19 (page 44 [1]) there is given a schematic of a controlled object with a linearly acting controller (Fig. 1), and in Fig. 29 (page 62 [1]), there is given a schematic for a controlled object with a nonlinearly acting controller without feedback (Fig. 2).

On the basis of the schematic of Fig. 1, M. V. Semenov draws the conclusion that ". . . a linearly acting automatic controller is comprised only of a sensing element."

It should follow then from Fig. 2 that a nonlinearly acting controller consists of a sensing element and a servomechanism. Thus, using M. V. Semenov's text, one may enunciate the following:

1. An automatic control system consists of three elements: regulated object, regulating organ and auto-

*** The editor considers very important the development of criticism by the readers in the direction indicated in this article.**

matic controller. The automatic controller and the regulating organ comprise the controlling device, i.e., besides the concept of automatic controller there is now introduced the concept of controlling device.

2. The regulating organ is not an element of the automatic controller.

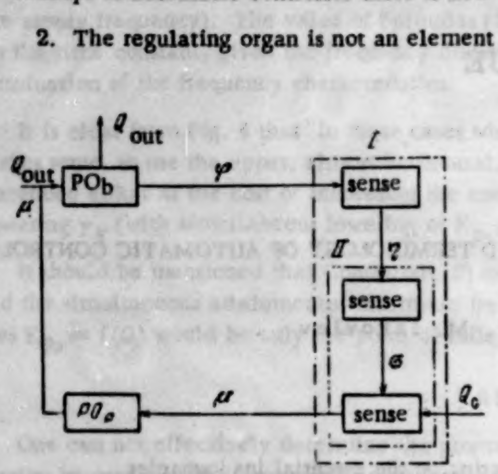


Fig. 2. I is the automatic controller, II is the executive mechanism.

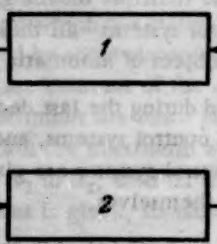


Fig. 3. 1 is the regulator, 2 is the controlled object.

M. A. Aizerman [2] says on page 23: "A system of straight-line control consists of two elements, connected in a closed loop. One of the elements is the controlled object and the second is the controller . . . . In the simplest cases, a nonlinear control system consists of three elements: the controlled object, the sensing element, reacting to deviations of the controlled quantity from its required value, and a servomotor, controlled by the sensing element." In Fig. 7b [2] is shown the interaction schematic for a linear control system, and in Fig. 9b [2], the schematic for a non linear control system (Fig. 3 and 4).

On page 22 [2] it is asserted that: "In a linear control system, the controller has one degree of freedom. It consists of one element which reacts to variations of the controlled quantity and immediately affects the regulating organ of the regulated object." As is clear from Fig. 3 and 4, M. A. Aizerman does not show in his schematic the action of the regulating organ as one of the elements of the automatic control system, as does M. V. Semenov.

In his book [3], E. P. Popov says on page 20: " A system of automatic control consists of a controlled object and a controller . . . By the <u>regulator</u> is meant the automatic device <u>perceiving</u> undesirable deviations of the controlled quantity <u>x</u> and so <u>acting</u> on the controlled object that these deviations are removed . . . . For implementing its functions, the regulator first must have at its input a <u>sensing element</u> for perceiving the undesired variations of the controlled quantity and, second, the regulator must have at its output a <u>controlling organ</u>, exerting on the controlled object a definite influence toward expunging the generated undesirable deviation of the controlled quantity. In addition to the sensing element (at the input) and the controlling organ (at the output), the controller may have various intermediate devices." Corresponding to this is given, in Fig. 8 (page 21 [3]), a structural schematic (Fig. 5).

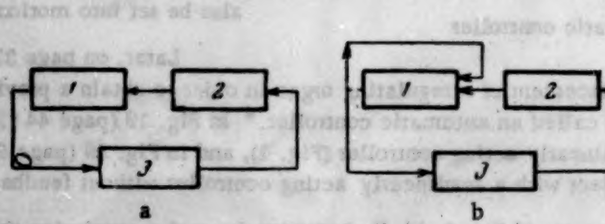


Fig. 4. 1 is the servomotor, 2 is the sensing element, 3 is the controlled object.

In another book by E. P. Popov [4], Fig. 12 (page 23 [4]) shows the block schematic of an automatic control system with linear control (Fig. 6). Thus, according to E. P. Popov, a linear automatic controller consists of a sensing element and a controlling organ, but a nonlinear controller, in addition to the aforementioned elements, also includes intermediate devices. It is clear from just the three texts considered that there does not exist unanimous agreement as to the elements comprising automatic control systems and automatic controllers, and that the block schematics given in these texts appear quite diverse. After reading the above-mentioned texts, one in-

voluntarily raises the question: to which portion of the controlling system should the controlling organ be referred? Is it an element of the automatic controller or of the controlled object, or is it necessary that it be considered an individual element of the system of automatic control? At firt glance, this question might appear trivial. However, we feel that the given question is a cardinal one, for the following reasons:

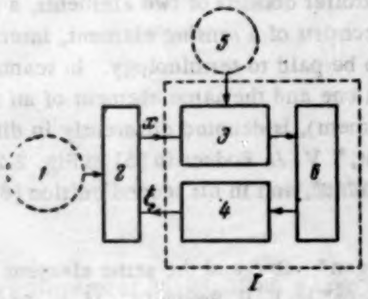


Fig. 5. I is the controller; 1 is the disturbing influence f(t), 2 is the controlled object, 3 is the sensing element, 4 is the controlling organ, 5 is the adjustment y, and 6 is the set of intermediate devices.

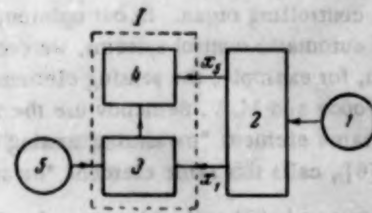


Fig. 6. I is the automatic controller; 1 is the disturbing influence, 2 the controlled object, 3 the sensing element, 4 the controlling organ, 5 the adjustment.

1) In studying automatic control systems, using the literature sources, one gets different opinions as to the basic elements of automatic control systems and, in some cases, as to the elements entering into the makeup of the automatic controller and the controlled object;

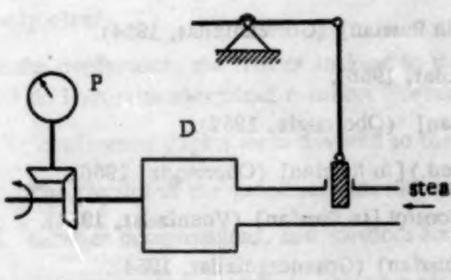


Fig. 7.

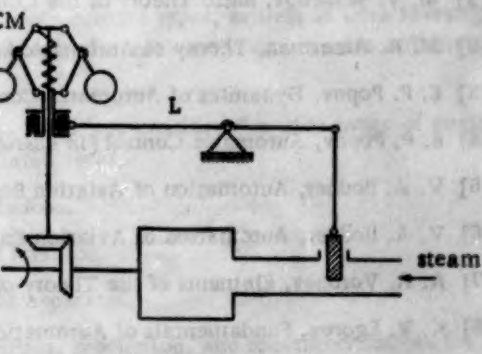


Fig. 8.

2) The automatic regulators actually used in technology fall into two variants: one has built in both sensing element and controlling organ, and the other has only the sensing element.

For example, the linear action, angular stress regulator, employed in aircraft technology, consists of a sensing element and a controlling organ, implemented in one box, while the regulator of the aircraft cabin temperature, a nonlinear one, consists only, essentially,of a sensing element and intermediate devices. However, the first and the second both bear the name of automatic controllers ("avtomaticheskogo reguliatora).

In our opinion, it is necessary to consider as automatic controllers devices consisting of the following elements: a sensing element, intermediate devices and a controlling organ (reguliruiushchii organ). In support of this assertion, we consider a system for maintaining rotational stability of a heat engine.

With manual regulation (Fig. 7) a human observes the rotation of the engine by means of the indicator P, compares the actual rotation with the desired value, and by means of a valve Z increases or decreases the supply of energy to the engine, D. This is not automatic regulation, but the human, together with the enumerated devices, does execute the role of automatic controller. To make the system automatic, it is necessary to set up a centrifugal mechanism and connect this to the valve.

A human alone does not play the role of an automatic controller. He needs supplementary elements: a counter of revolutions and a valve. In automatic regulation by the centrifugal mechanism CM (Fig. 8), one part of the automatic controller does not appear. For this, it is necessary to have a connection to the valve via a kinematic linkage, L.

Consequently, it is correct to state that a linear automatic controller consists of two elements, a sensing element and a controlling organ, but a nonlinear automatic controller consists of a sensing element, intermediate devices and a controlling organ. In our opinion, attention should also be paid to terminology. In scanning block schematics of automatic control systems, we convinced ourselves that one and the same element of an automatic control system, for example, the sensing element (chuvstvitel'nyi element), is denoted differently in different books. E. P. Popov and M. V. Semenov use the term "sensing element," V. A. Bodner in [5] in Fig. 2.2 (page 29 [5]) calls this same element "measuring sensing element of the controller", and in his second edition [6], in Fig. II.3 (page 31 [6]), calls this same element "measuring device".

A. A. Voronov in [7] calls this same element the "measuring organ". One and the same element of the system is called "controlling organ" by A. A. Voronov [7], "controlling element" by E. P. Povov [3], M. V. Semenov [1] and K. V. Egorov [8], and "amplifier" by V. A. Bodner [5]. The driving device is called "servomotor" by V. A. Bodner and M. V. Semenov, "power element" by K. V. Egorov, and "drive" by E. P. Popov. It seems to us that enough has beeen said in this connection to warrant drawing certain conclusions. It is time to develop unified concepts and definitions, and to adhere to one terminology in questions of automatic control. The question is placed open to general discussion.

It would seem that if the authors of the texts mentioned were to enter into this discussion, and whole-heartedly, the problem would be solved.

LITERATURE CITED

[1] M. V. Semenov, Basic Theory of the Control of Power Machines [in Russian] (Trudy LK VVIA, 1950).

[2] M. A. Aizerman, Theory of Automatic Engine Control [in Russian] (Gostekhizdat, 1952).

[3] E. P. Popov, Dynamics of Automatic Control Systems [in Russian] (Gostekhizdat, 1954).

[4] E. P. Popov, Automatic Control [in Russian] (Gostekhizdat, 1956).

[5] V. A. Bodner, Automation of Aviation Engines [in Russian] (Oborongiz, 1952).

[6] V. A. Bodner, Automation of Aviation Engines (second ed.) [in Russian] (Oborongiz, 1956).

[7] A. A. Voronov, Elements of the Theory of Automatic Control [in Russian] (Voenizdat, 1954).

[8] K. V. Egorov, Fundamentals of Automatic Control [in Russian] (Gosenergoizdat, 1954).

CHRONICLE

CONFERENCE ON ELECTRICAL CONTACTS

During November 26-28, 1956, a conference on electrical contacts was held in Moscow. The conference was organized by the IAT (Institute for Automation and Remote Control) of the AN SSSR, in conjunction with various research institutions. One hundred seventy people from 53 organizations participated: 33 research institues and 20 plants turning out electrical apparatus and contacts. At the conference 25 papers were heard, after which a 22-man panel discussion was held. At its conclusion, the conference adopted several resolutions.

The aims of the conference were to clarify the present state of research work in the domain of contacts and the phenomena related to contact functioning, the investigations into increasing the useful life of contacts and determining the reliability of their use, and the questions of production and utilization of contact materials.

Electrical contacts figure among the most wide-spread and critical of the elements in electrical designs. On contact properties depend, in large measure, the reliability and length of useful life of electrical apparatus. It is in this connection that, recently, there has appeared a large quantity of work in the determination of useful life and domain of rational employment of contact materials and various contact types, as well as work investigating the phenomena connected with electrical contacting functioning, in so far as those processes were hitherto not completely clear.

In the conference, the first of its kind in the Soviet Union, there was considered a wide range of questions, having to do both with electrical contacts directly and with related areas.

The conference papers were devoted to the following questions.

1. The physics of the processes entering into the contact function.

2. Contact computations, and methods for testing contact apparatus.

3. Current contact materials, their characteristics, properties, production, and recommendations for their usage under various conditions.

In the papers devoted to the physics of the processes entering into contact functioning, there were explained several endeavors in the investigation of spark discharge and the electric spark machining of metals.

S. L. Mandel'shtam, V. P. Shabanskii and N. K. Sukhodrev (Moscow) showed that spark discharges are characterized by high energy density produced at the electrodes, whereupon the leakage from the cathode occurs explosively.

L. S. Palatnik (Kharkov State University) showed that the fundamental role in the transfer phenomenon in electric spark machining is played by the physico-chemical nature of the electrodes and their environment, and stated criteria for the transfer and the mutual interaction of the electrodes.

B. N. Zolotykh (Central Research Laboratory for Electric Spark Machining of Materials, AN SSSR) showed that the amount of erosion is related to the heat constant of the material, and proportional to the energy supplied in the pulses.

N. V. Afanas'ev (White Russian Polytechnic Institute) compared the physical constants of materials with their erosion stability, and attempted to establish certain connections between the magnitude of the erosion and the heat of evaporation. In this group should also be included the papers of I. V. Kragel'skii (Institut Mashinovedeniia AN SSSR), who demonstrated both theoretically and on models the actual dependence of the area of the contact surface

on the properties of the surface; the paper of A. I. Pugin (Metallurgical Institute AN SSSR), investigating the character of the initial formation of the contact, the potential difference on the contact, and the temperature in the neighborhood of the contact during a welding process; and the paper of M. A. Razumikhin (Moscow) which presented methods of coping with bridge and spark erosion, and on the influence of the environmental pressure on the functioning of a contact.

Among the papers dealing with the second group of questions should be included the paper of B. S. Sotskov (IAT AN SSSR) which expounded an approximate method for computing relay contacts for different conditions of use, taking into account their erosion and the probability of their breakdown as a function of their length of service, the paper of O. B. Bron (Leningrad), which gave an analysis of the conditions of functioning of contacts in contactors and in automatic cut-outs, the dimensions and parameters of commutating links with and without arc quenching, and the papers of G. O. Feiler (Moscow) and V. I. Fiks and M. A. Meerovich (Plant ATE-1 Ministry of the Automotive Industry), presenting the results of industrial tests of different contact materials, and the paper of A. V. Gordon (Moscow), reporting on the requirements presented by contacts for aircraft contactors, and stating methods for accelerating contact testing.

The third group of questions was illuminated by a large number of papers dealing with new contact materials, with certain peculiarities of previously known materials, and with the replacement by these of the precious metals.

A. B. Al'tman, I. P. Melashenko and E. S. Bustrova (Moscow) reported on existing methods of producing metallo-ceramic contacts, and showed that during their operation the surface layer of metallo-ceramic contacts was changed in structure and in phase composition.

I. N. Frantsevich and O. K. Teodorovich (Metallurgical Institute AN USSR) state that the most economical and general method of producing metallo-ceramic contacts is the impregnation method, and that for obtaining material with the very best characteristics it is necessary to compress the composition material.

I. E. Dekabrun (IAT AN SSSR) explained the results of work in the determination of certain characteristics of metallo-ceramic contacts, and recommended certain usages of such contacts in the domain of weak currents.

V. V. Usov and E. M. Murav'eva (Moscow) reported on contact materials obtained by internal oxidation of certain alloys, the properties of which approximate to the properties of metallo-ceramic contacts, but the technology of preparing which are significantly simpler and cheaper.

A. B. Al'tman and E. S. Bustrova (Moscow) reported on the possibility of using aging metallurgical alloys for contacts, and showed that their stability under wear is very high.

V. V. Usov and M. D. Povolotskaia (Moscow) presented data on the reasons for corrosion of tungsten contacts; E. A. Shumskaia (NII Avtopriborov Ministerstva avtomobil'noi promyshlennosti) showed that the erosion of tungsten contacts depends on the deformation of their atomic lattices.

A. A. Rudnitskii (Metallurgical Institute AN SSSR) presented data on the thermoelectromotive force of a series of precious metals and their alloys, which plays a large part in the behavior of low current contacts.

E. S. Kirillova (Moscow) compared the characteristics of palladium-silver and other alloys, determined with the object of using these alloys for sliding contacts with very small pressures.

E. K. Iuferova (Moscow) made recommendations on the use of a series of new materials for sliding contacts in synchronous transmission systems.

V. A. Mitiushev and V. V. Kozlov (Moscow) made a survey of the research work being done in industrial plants, and reported on the state of contact and contact material production in these plants.

The discussions brought out the great value of the conference for the wide interchange of experience and the dissemination of results obtained. Moreover, a series of remarks was made in the papers, reporting on a host of new labors, carried out in the area of contact research, and a number of assertions were aired.

Among the resolutions adopted by the conference should be mentioned the necessity for periodic sessions (one every two or three years) of the conference on electric contacts, and also the necessity of coordinating research work in the investigation of processes occuring in contacts, in the investigation of useful life, in the development of new contact materials, and in the establishement of standard methods for testing contacts.

I. E. Dekabrun